National Center
for Toxicogenomics

SUMMARY OF THE

SYMPOSIUM ON GENE EXPRESSION AND PROTEOMICS IN ENVIRONMENTAL HEALTH RESEARCH

DECEMBER 3-4, 2001
NATIONAL INSTITUTES OF HEALTH
BETHESDA, MARYLAND

Classifying Toxicants

Biology Is An Information Science

Toxic Torts

Dissecting Gene Regulatory Networks

Systems Biology

Computer-Intensive Algorithms

Classifying Cancer Subtypes

Phenotypic Anchoring

Summary of the

SYMPOSIUM ON GENE EXPRESSION AND PROTEOMICS IN ENVIRONMENTAL HEALTH RESEARCH

presented by the
National Center for Toxicogenomics
National Institute of Environmental Health Sciences
National Institutes of Health

at the Natcher Center Auditorium
National Institutes of Health
Bethesda, Maryland

December 3-4, 2001

Symposium Co-Chairs:
Leona D. Samson, Massachusetts Institute of Technology
James K. Selkirk, National Institute of Environmental Health Sciences, NIH
Raymond W. Tennant, National Institute of Environmental Health Sciences, NIH

Adminstrative Report prepared by

Miriam Sander
Page One Editorial Services
Durham, North Carolina

Table of Contents

Foreword

On December 3-4, 2001, the National Center for Toxicogenomics (NCT) and the National Institute of Environmental Health Sciences, National Institutes of Health, sponsored the Symposium on Gene Expression and Proteomics in Environmental Health Research at the National Institutes of Health, Bethesda, Maryland. This Symposium was held to assess recent developments in the field of toxicogenomics, to promote exchange of information and ideas, and to promote future progress and achievements in environmental health research.

This report summarizes the main points of the presentations made at the Symposium. It is provided as a guide to the activities and interests of the NCT, and as a reference for scientists who attended the Symposium. The NCT hopes that this report will also stimulate the interest of scientists who did not attend the Symposium and help to promote future progress in environmental health research.

Kenneth Olden
Director
National Institute of Environmental Health Sciences
National Institutes of Health

About the National Center for Toxicogenomics

The National Institute of Environmental Health Sciences (NIEHS) established the National Center for Toxicogenomics (NCT) in June 2000. Toxicogenomics is an emerging scientific field that combines studies of genetics, genome-wide mRNA expression, cell and tissue-wide protein expression, and bioinformatics to understand the roles of gene-environment interactions in disease. The NCT was created to catalyze application of toxicogenomics to improve human health.

The goals of the NCT are: 1) to develop and apply gene expression and proteomics technology to study the biological effects of chemicals and drugs; 2) to support intramural and extramural research to define the effects of environmental agents on gene expression; and 3) to develop a national reference and relational database on "Chemical Exposure in Biological Systems" (CEBS) that will serve as a resource in the fields of toxicology and environmental health.

To further its goals, the NCT, NIEHS and NIH organized the **Symposium on Gene Expression and Proteomics in Environmental Health Research**. The symposium was held December 3-4, 2001 at the Natcher Center at the National Institutes of Health in Bethesda, Maryland. More than 400 participants attended the 2 day symposium on the state-of-the-art in Toxicogenomics.

Symposium Overview

This symposium opened with a keynote address by **Leroy E. Hood** (The Institute for Systems Biology) and a Special Lecture in Environmental Genomics by **Michael Karin** (University of California, San Diego). Hood's talk took a broad view of modern biology as an information science in the post-genomic era. Karin's talk presented a detailed and informative discussion of cell signaling pathways that play important roles in environmental diseases. In the five symposium sessions that followed, session chairs and speakers captured the state-of-the-art in Environmental Genomics, Gene Expression, Proteomics, Biomarkers, Toxicogenomics, and the Ethical Legal and Social Issues of the post-genomic era. The symposium ended with a review of the highlights of the symposium presented by **Samuel H. Wilson** (Deputy Director, National Institute of Environmental Health Sciences, NIH).

In his closing remarks, Wilson emphasized the importance of Hood's concept of systems biology in the post-genomic era. The task of understanding any biological system is enormous, and we don't yet understand even the simplest single cell organism. Hood pointed out the value of model organisms to help decipher and analyze biological systems. Wilson also emphasized the importance of Hood's systems biology approach, in which elements of a biological system are defined, a system model is formulated, and the model is tested by perturbing and re-analyzing the system in an iterative manner.

Wilson pointed out that we will use discovery science to define the elements of biological systems. New technology will play an important role in this process, as the rate at which data can be collected, stored and analyzed continues to increase exponentially. Toxicogenomics is not a static field; new methods and tools will have to be embraced and understood as they emerge. New visual and computational tools will be needed as well. For example, structural biological analysis of macromolecules and macromolecular complexes is an important area of research in which the data can not be visualized in 2-dimensionsal space. Structural biological problems and other highly complex biological problems now must be visualized in 3-dimensional or higher dimensional space; lower dimensional analysis is insufficient to represent the structural and informational complexity inherent in biological systems.

The importance of understanding cell signaling and protein networks was highlighted throughout this symposium, but especially in the talk presented by Michael Karin. Karin showed the importance and public health relevance of stress-activated protein kinases and mitogen-activated protein kinases. Karin pointed out that we must understand these pathways in order to understand public health issues such as arsenite toxicity and type 2 diabetes. Hood, Bing Ren and Roger Ulrich also spoke to the potential of toxicogenomics to decipher protein networks and the complex protein pathways that contribute to pathophysiology.

Wilson suggested that the most important message to take away from this symposium is an appreciation of the "success stories in reduction to practice" in this new field of toxicogenomics. Wilson catalogued the following examples presented during the symposium: Chris Bradfield, Richard Paules and Roger Ulrich classified unknown toxicants/drugs using gene expression microarray; David Duggan classified breast cancer subtypes using gene expression microarray; Edwin Clark classified stages of ovarian cancer using gene expression microarray; Emmanuel Petracoin classified stages of ovarian cancer using proteomic markers; Richard Paules classified liver toxicity endpoints using gene expression microarray. These results definitively show that toxicogenomics can used to classify biological endpoints of toxicity and pathology. The query is no longer whether toxicogenomics will be a useful tool in the fields of toxicology and pharmacology. In addition, Wilson concluded that "it is no longer in question whether toxicogenomics can live up to its promise to revolutionize the field of toxicology. This assertion is on firm ground."

In looking to the future of toxicogenomics, Wilson listed several important areas of focus. 1) New technology should be anticipated and used to facilitate progress and overcome existing bottlenecks; 2) Systems biology will be essential for trying to integrate the fields of biology and pathology and understand pathogenic mechanisms of complex diseases; 3) Uniformity and quality control of toxicogenomics methods needs to improve; to this end, best practices, standards and guidelines for study design should be developed and widely disseminated in the research community; 4) A robust, well annotated public database will also be extremely important to the future of toxicogenomics; 5) And finally, as demonstrated by the interest generated by the ELSI session at this symposium, ELSI and education concerning ELSI issues should be given high priority as toxicogenomics begins to have greater impact on public health issues.

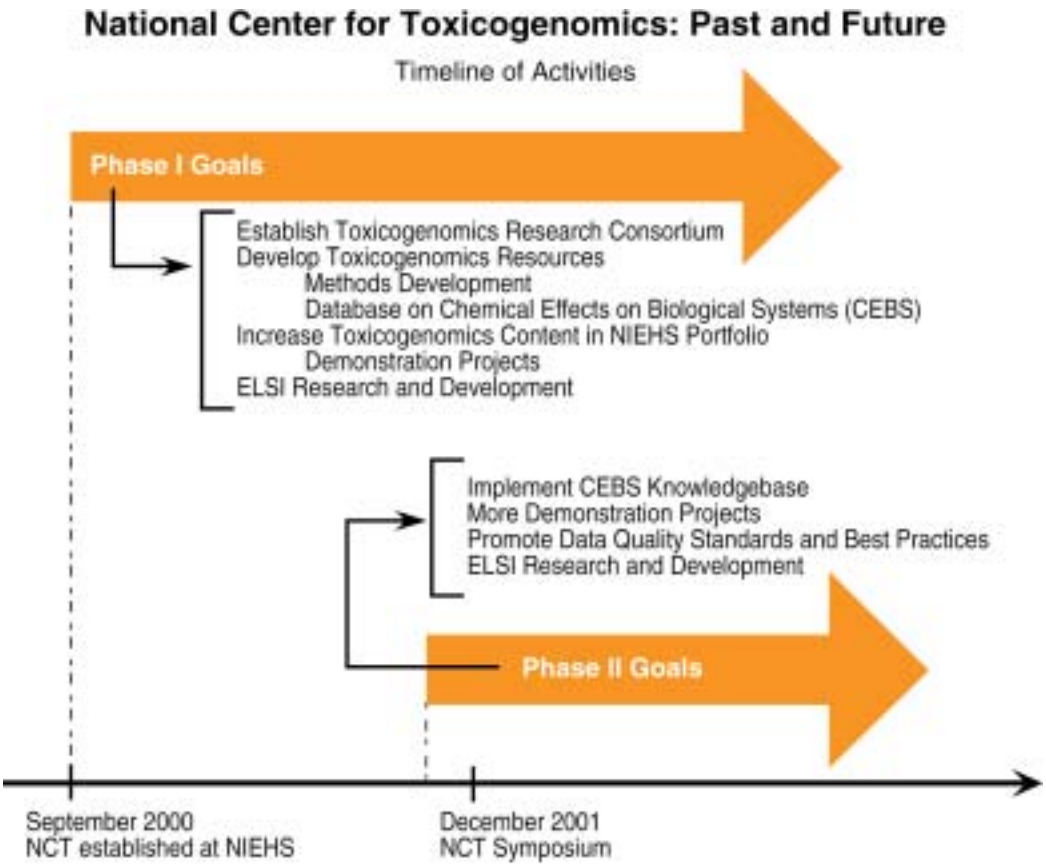

Figure 1

The NCT has made excellent progress in achieving its Phase I goals: establishment of a Toxicogenomics Research Consortium; incorporating toxicogenomics in its portfolio; and developing resources to support toxicogenomics research, including the CEBS database (Figure 1). Work on Phase I goals will continue as the NCT moves on to other key goals during Phase II. More demonstration projects will be undertaken and best practices will be developed and disseminated. In addition, the NCT is committed to promoting broad access to toxicogenomics resources within the entire extramural research community. This symposium is a testament to the progress made by the NCT in its first two years, and the talks presented here are a good representation of the current state-of –the-art of toxicogenomics.

Introductory Remarks

Kenneth Olden (Director, National Institute of Environmental Health Sciences, NIH) opened the symposium by pointing out that the fields of toxicology and environmental health sciences have evolved slowly up to now, because the tools available to scientists working in these areas were relatively unsophisticated. These tools supported slow progress because they were designed to ask questions about one chemical, one endpoint or a small number of genes.

The completion of a draft sequence of the human genome has brought both new opportunities and new challenges to toxicologists and environmental health scientists. One challenge is to use the human genome sequence to understand the genetic and biological basis of complex biological traits and diseases such as cancer, diabetes, Alzheimer's Disease and Parkinson's Disease. Another challenge is to meet the increased demand for toxicological information and progress in understanding toxic responses. Advances in combinatorial chemistry and molecular biology have dramatically accelerated the rate of drug discovery and the rate at which populations are exposed to new drugs and chemicals. This trend increases the burden of exposure in the population, and makes it critical that we rapidly increase our understanding of the consequences of such exposure.

The field of toxicology could not have risen to this challenge using only the less inefficient technologies of the past several decades. The field needed more informative, more cost-effective and more efficient tools and technologies to answer increasingly complex questions and an increased demand for research progress. We now have these technologies and are in a position to carry out toxicological inquires on a global scale. The post-genomic era has given researchers the capacity to carry out global analysis of mRNA and protein expression using gene expression DNA microarrays and proteomics. It is now possible to examine multiple endpoints and, for the first time, to study the effects of complex mixtures. Proteomics, toxicogenomics and metabolomics will help the scientific and medical community understand complex biological pathways and complex toxicological responses in biological systems.

In the post-genomic era, we are also beginning to have insight into human genetic susceptibility to disease. Databases are being developed that catalogue human DNA variation and map the associations between DNA polymorphism and human disease susceptibility. These technological advances bring great promise to the field of toxicology and make this an extremely exciting time to be involved in environmental health research. In the near future, it is likely that toxicogenomics will have a significant impact: this emerging field has the potential to dramatically increase our understanding of complex toxicological problems and enhance our ability to improve and protect human health.

Keynote Address

In his keynote address, **Leroy E. Hood** (The Institute for Systems Biology) emphasized the importance of recent paradigm shifts in the scientific community which are the result of the Human Genome Project and the completion of a draft human genome sequence. The first paradigm shift concerns the advent and feasibility of systems approaches in biology (*i.e.*, systems biology) and the second paradigm shift is a dramatically altered view of predictive and preventive medicine. The Human Genome Project led to these paradigm shifts at least in part because the sequence of the human genome provides a "genetic parts list" for scientists and health professionals to use in systematic analyses of the human organism. Hood also strongly emphasized that biology is now an information science, and he explained the significance and implications of the information age of biology. Biology has changed dramatically in the past 10 years: biological methods have changed, the quality and quantity of biological data has increased rapidly, and equally importantly, there has been significant change in the way biology is perceived both within and outside the scientific community.

The Systems Biology Approach

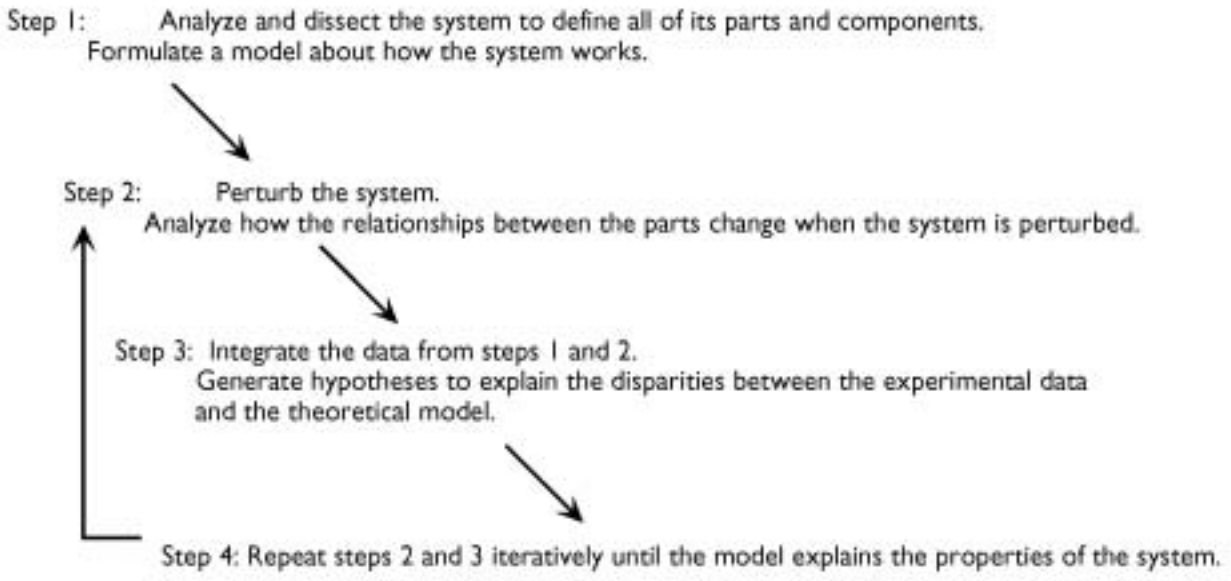

Figure 2

Hood used the analogy of a car to explain the systems biology approach (Figure 2). To understand a complex machine like an automobile, the first step is to describe or define each of the automobile's parts and to formulate a model about how the car functions as a system. The second step is to perturb the system and to analyze how the relationships between the parts change as the system is challenged. The third step is to integrate the data obtained in steps 1 and 2 and to generate hypotheses that explain the disparities between the experimental data and the theoretical model. Importantly, this process must be carried out iteratively, repeating steps 2 and 3 many times. This iterative systematic process is equally applicable for a system like a car and for a system like a bacterium, mouse or human body. It is now applicable to biology precisely because biology has become an information science.

Five characteristics of biological information were enumerated by Hood as follows:
1) The human genome sequence is "knowable" because it is encoded using a four base digital source code, and thus it is digital information.

2) The human genome provides two kinds of information: information about human genes and information about the regulatory networks that control the activity of and interactions between human genes.
3) Biological information is historical.
4) Biological information is hierarchical.
5) Biological information comes from three sources: genes, the environment and stochastic processes.

The keys to understanding evolution, physiology and development lie in the gene regulatory networks that control the activity of complex biological machines and pathways. Hood suggested that comparative genomics and the study of model organisms are the two most important methods that can be used to deconvolute these complex relationships and networks. Comparative genomics brings focus to regulatory mechanisms and networks that are unique to each organism. For example, the genomes of apes, chimps and humans are very similar in their gene coding regions; however, there are very large differences in the functions and capabilities of the brains of these three distinct species. These species differences can not be accounted for by single nucleotide polymorphisms in gene coding regions. Thus, the differences are most likely due to gene regulation and genetic interactions that are unique to each species.

Hood illustrated the hierarchical and historical nature of biological information by describing a developmental mutant of the model organism *Drosophila.* The *Drosophila* mutant fly has two sets of wings, whereas a normal fly has one set of wings. The mutation that causes this phenotype is in the coding region of a developmentally regulated transcription factor. The mutant transcription factor acts within the context of the normal developmental program of the fly, creating a set of wings, but doing so either at the wrong time or in the wrong place, or both. It is clear that the function of both the normal and mutant
transcription factor is constrained by and reflects the hierarchy of the pathway in which it normally functions; thus, this example demonstrates the hierarchical nature of biological information.

Although normal fruit flies have 2 wings and abnormal fruit flies have 4 wings, four-wingedness is the normal morphology for an evolutionary precursor of *Drosophila.* Thus, the four-winged *Drosophila* mutant mimics the form of the evolutionary ancestor from which it evolved. This example also demonstrate that a single amino acid change in a transcription factor can recreate a historically related organism. Biological information is therefore historical in nature.

It is fairly self-evident that genes and the environment are sources of biological information. It is a bit less obvious that stochastic processes also inform and determine the biological characteristics of an organism. Hood demonstrated this fact by considering the example of a pair of identical twins. These individuals have identical genes, and the identity of those genes is responsible for the physical and biological similarity between the twins. However, because the twins experience different environments, over time the two individuals accumulate unique attributes that reflect environmental influences to which they are exposed. The fingerprints of the twins are non-identical and it is highly unlikely that the unique fingerprint of each twin is determined by environmental factors. In contrast, it is more likely that the unique fingerprint of each twin represents the effects of random or stochastic events that influence a unique and non-reproducible pattern of skin growth during development. Hood believes that the role of random events in biological systems has until now been poorly recognized, but that stochastic processes should be considered in analyzing a biological system.

The genomic and post-genomic eras are characterized by the use of high throughput methods which generate massive amounts of data. Moore's Law, an axiom of the computer technology industry, predicts that computing capacity will double every 18 months, and the same general trend is apparent in the increase in capacity of high throughput molecular biological methods. Hood predicted that it will be possible to sequence an entire genome in one day within 10 years. Hood also described newly

developed high throughput methods that are revolutionizing the scientific world, dramatically increasing the rate at which the genome, transcriptome and proteome can be analyzed. Along with this rapid increase in data output comes the need to analyze, store, integrate, visualize, model and distribute this newly acquired biological information. Hood emphasized that one of the biggest challenges of the future lies in this area. How can appropriate and sufficiently powerful computational methods and models be developed to integrate and analyze biological information? This is an extremely difficult and urgent task.

Systems biological approaches and the use of model systems are essential in the age of biology as an information science. Model organisms are like a Rosetta Stone of biological information, allowing us to translate the informational pathways intrinsic to different biological organisms, in the same manner as the Rosetta Stone allowed us to break the code of ancient foreign languages. Systematic and global approaches can be used to understand model organisms, leading from comprehensive understanding of a single cell organism such as yeast, to comprehensive understanding of a mouse or a human being.

Hierarchical biological information exists at many different levels, from DNA and RNA, to proteins and informational pathways, to populations and ecologies. Systems approaches may provide a way to integrate all levels of biological information and understand how complex attributes at the top levels of the hierarchy relate back to the source code or to intermediate elements within the hierarchical system. Thus, systems biology can be used to understand the continuity thru the hierarchy of a coherent system of biological information.

Special Lecture in Environmental Genomics

Michael Karin (University of California, San Diego) discussed the combined effect of genes and environmental factors in determining the phenotype and the relative health and/or disease status of an organism. Karin is particularly interested in the role of environmental factors in common polygenic diseases such as asthma, type 2 diabetes, irritable bowel syndrome, rheumatoid arthritis, atherosclerosis and cancer. Karin believes that these diseases, which impact a large proportion of many human populations, are appropriately viewed as "environmental diseases," because environmental factors play a significant role in their etiology. Karin also pointed out that these are complex diseases, in which a single genotype associated with the disease can cause multiple phenotypes. This reflects the fact that these diseases are caused by multiple genetic and environmental factors that interact with each other, modulate each others effects, and have different levels of penetrance.

Signaling pathways are an important mechanism by which environmental factors modulate gene expression. One such signaling pathway involves stress activated protein kinases including c-jun N-terminal kinase (JNK), p28 and I-kappaB kinase (IKK). Karin described two examples of how environmental factors interact with this signaling pathway. The first example focuses on the activity of NF--κβ as a an essential negative regulator of IKK-induced apoptosis. The second example demonstrates that glucose homeostasis is disrupted by activation of JNK1 in response to excess dietary free fatty acids (FFA). The latter example has implications for prevention and treatment of type 2 diabetes.

The NF--κβ signaling pathway is activated by many different stimuli including microbial pathogens, byproducts of microbial infection, chemokines and other proinflammatory stimuli. The pathway is activated through the superfamily of tumor necrosis factor (TNF) α and β receptors, the IL-1 receptor superfamily and toll-like receptors. These receptors activate IKK, which leads to NF--κβ-mediated changes in gene expression. Gene targets of NF--κβ include inflammatory enzymes, adhesion molecules and chemokines. However, NF--κβ is also a major negative regulator of apoptosis, although the mechanism by which NF--κβ down-regulates apoptosis is unknown. NF--κβ plays a role in tumor development by its direct and indirect effects on cell proliferation. Activation of NF--κβ requires a functional IKK kinase complex.

The IKK kinase complex is composed of α and β catalytic subunits and a regulatory γ. subunit. IKK α and β share a homologous "activation loop" that includes two serine phosphorylation sites which are phosphorylated in the active form of IKK. IKK is activated constitutively when the IKK β serine residues are mutated. The activation loop in IKK β also includes a cysteine that plays a role in a negative feedback loop of IKK. Several mechanisms activate this negative feedback loop: IKK β is inhibited by prostaglandin J2 (produced by COX-2) when the prostaglandin interacts with this regulatory cysteine residue; in addition, arsenite interacts with this cysteine residue and downregulates IKK activity. In both cases, inhibition of IKK can stimulate apoptosis. This mechanism explains the efficacy of arsenic in potentiating the ability of chemotherapeutic agents to kill tumor cells.

Karin created transgenic IKK β knockout mice and mice with tissue-specific deletions of IKK β. The phenotype of these mice demonstrate that IKK β and the NF--κβ signaling pathway play an essential role in negative regulation of apoptosis. The IKK β knockout mouse dies during mid-gestation due to massive TNF α-mediated liver necrosis. Similar effects are observed when IKK β is inactivated in a tissue specific manner. Karin suggests that many hepatotoxins may also cause liver necrosis by interfering with activation of IKK and/or NF--κβ.

In the second part of his talk, Karin described the role of JNK-mediated signaling on insulin responsiveness in mouse models for type 2 diabetes. This study focused specifically on environmental

stimuli that modulate the activity of insulin receptor substrate 1 (IRS-1) by phosphorylation. IRS-1 is activated by tyrosine phosphorylation. and the active form of IRS-1 is required for insulin responsiveness. In patients with type 2 diabetes, FFA and TNF α play a role in the development of insulin resistance. One mechanism for this effect involves negative regulation of IRS-1 via serine phosphorylation. Karin studied this pathway with the goal of identifying the kinase cascade responsible for FFA-induced insulin resistance.

There are several useful mouse models for mice diet-induced obesity (obese mice) and type 2 diabetes (leptin-deficient ob/ob mice). Karin examined physiological parameters, the activity of JNK1/2, IRS-1 phosphorylation, insulin sensitivity and glucose tolerance in these mice. Similar experiments were carried out with JNK1- or JNK2-deficient mice. Ob/ob mice upregulate JNK in liver, fat and muscle. In the model for diet-induced obesity, mice also upregulate JNK when fed a high fat diet. However, when JNK1 is inactivated by mutation, these animals have less subcutaneous fat, smaller adipocytes, increased insulin sensitivity, lower body weight and normal muscle mass. Deletion of JNK2 does not alleviate diet-induced obesity in these animals. Karin concludes that activation of JNK1 plays a significant role in diet-induced obesity.

One mechanism for JNK1-mediated diet-induced obesity could involve down-regulation of IRS-1. Karin tested this possibility, and demonstrated that IRS-1 tyrosine phosphorylation increases in a JNK1 mutant. Thus, JNK1 may down-regulate IRS-1 by increasing serine phosphorylation of this protein in response to dietary FFA.

Karin emphasized that the JNK1 pathway may play a critical role in glucose homeostasis and diet-induced obesity. Thus, JNK1 is key link in signaling in response to environmental stimuli including dietary FFA. It is possible that JNK1 inhibitors could be developed as therapeutic agents to prevent or treat type 2 diabetes, one of the most common environmental diseases in the human population.

Reports of Conference Sessions

Session I: Environmental Genomics

Session I Speakers:
Chair: Daniel W. Nebert, University of Cincinnati Medical Center
Chris A. Bradfield, McArdle Laboratory for Cancer Research
Robert B. Weiss, University of Utah
Deborah A. Nickerson, University of Washington School of Medicine

Session I Introduction and Highlights:
Daniel Nebert chaired the session on environmental genomics, which included one talk on global expression analysis as a means to classify and understand toxicants, and two talks on human DNA polymorphism. Nebert introduced the topic of human DNA polymorphism by defining a polymorphic DNA variant as a sequence variation with a frequency of 0.01 or greater. A rare variant is represented at a frequency of 0.01 or less. At least 3 million human single nucleotide polymorphisms (SNPs) have already been catalogued, and it is likely that there are as many as 11 million SNPs (frequency of 0.01 or greater) per haploid genome. SNPs do not distribute randomly, but associate with one another on a single chromosome into a group called a haplotype. The characterization of human SNPs and haplotypes will become one of the most important tasks for the post-genomic era.

Highlights of this session included Chris Bradfield's use of cDNA microarray expression analysis to classify a series of liver hepatotoxins with 100% accuracy using a set of 12 diagnostic genes. Robert Weiss and Deborah Nickerson emphasized the importance of determining human haplotypes, especially when analyzing genotypes that contribute to complex diseases.

Session I Speaker Summaries:
Chris Bradfield carried out global expression analyses in mouse liver with the goal of classifying and understanding the mechanism of chemical toxicants. These two goals, classifying toxicants and defining their mechanism, are distinct objectives which are of general importance to toxicologists. Although classification of chemicals can be fairly straight-forward, it is often difficult to determine the mechanism of action of a toxicant. Bradfield focused on the Per-Arnt-Sim (PAS) protein superfamily which includes proteins that act as environmental sensors and play roles in adaptation. The best characterized PAS protein is the aryl hydrocarbon receptor (AhR).

AhR is a receptor for 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) and other related compounds. In the presence of its ligand, AhR induces expression of target genes, many of which protect the cell from xenobiotic compounds (*i.e.*, cytochrome p450s). However, activation of AhR is also associated with toxic effects including hepatomegaly, liver toxicity, developmental defects, tumors, thymic involution. The mechanism by which AhR mediates toxic effects is poorly understood. Bradfield established that these toxic effects require the transcriptional activity of AhR by studying a mouse that expresses a mutant AhR that is deficient only in transcriptional activation. He then decided to identify all transcriptional targets of the AhR, with the intent of determining which targets contribute to toxic effects of AhR.

One approach to global expression analysis is to create and analyze a comprehensive cDNA library. Bradfield favored this approach for identifying transcripts induced by TCDD, because it can identify tissue-specific transcripts, exposure-specific transcripts, paralogs, isoforms and splice variants. In addition, cDNA libraries can be used to directly measure expression frequency of a particular gene. cDNA libraries were created from the liver cells of mice treated with or without TCDD. Expression profiles were obtained for 2,500 mouse liver cDNAs. In animals treated with a single intraperitoneal dose of 10 mg/kg dioxin, the "inventory" of transcripts changed by approximately 20%; ≈ 10% of the

transcripts were expressed only in treated animals and ≈10% of the transcripts were expressed only in untreated animals. Specific gene changes included a 40-fold induction of cytochrome p450 cyp1a2. cDNA clones from mouse liver were also used to generate custom cDNA microarrays with up to 30,000 expressed liver sequences. Expression analyses using these microarrays demonstrated that approximately 20-25% of the genes were induced or suppressed 1.5-fold or more by exposure to TCDD.

Bradfield strongly believes that the complex response of the liver to TCDD and other toxicants depends on interactions between the different cell types in the liver. Thus, the next stage of this project will include analysis of dioxin-induced transcripts from endothelial, Kupfer and stellate cells in the liver. These cells will be isolated using a method base on cell type specific expression of green fluorescent protein.

Global expression analyses using DNA microarrays offers the possibility to streamline toxicology studies. One goal is to use this method to classify unknown toxicants. Bradfield demonstrated the feasibility of this approach using a training set of 5 classes of chemicals: dioxins, PCBs, peroxisome proliferators, hypoxia inducers and inflammatory compounds such as lipopolysaccharide and interleukin-6. He analyzed global expression in liver using a cDNA microarray chip with 1200 liver specific cDNAs. Many expression changes were noted and the profiles clustered according to chemical class. Using a large group of genes, classification was achieved with 80% accuracy. However, a small set of diagnostic genes is more desirable for distinguishing one class of chemical from another and could potentially be more accurate. Bradfield identified a set of 12 diagnostic genes that were 100% accurate in identifying unknown compounds that belong to these five chemical classes. Adding more genes to the diagnostic set only lowered classification accuracy.

Bradfield pointed out that the scalability of this approach is not known; nor is its general applicability. In addition, similar results may not be obtained when the approach is adapted for other chemical classes. Nevertheless, it is likely that microarray analysis will have broad applicability and success as a method to classify toxicants, thus facilitating many toxicology studies. It is less clear at present how great an impact microarray studies will have on mechanistic studies in toxicology, because these studies are intrinsically more difficult and complex. Success in mechanistic toxicology will require an integrated, multidimensional approach that is applied in a whole animal system.

Robert Weiss described SNP discovery using DNA from 90 ethnically diverse individuals obtained from National Human Genome Research Institute (NHGRI). This SNP discovery effort is being carried out by Weiss and the University of Utah Genome Center (UUGC) as part of the NIEHS Environmental Genome Project (EGP). The EGP is focusing on characterizing and resequencing approximately 500 genes thought to be relevant to environmental diseases. To date, 306 SNPs have been found in 19 genes covering a genomic region of 102 kb. The density of SNPs (nucleotide diversity) varies from a high of 1/164 bp to a low of 1/923 bp. Nucleotide diversity also varies in different regions of the same gene.

The SNPs identified by this project were compared with those in the GenBank SNP database, dbSNPs, which currently has approximately 3 million SNPs. Forty-eight of the 306 SNPs (16%) had been previously deposited in dbSNPs. In general, the SNPs found in both databases were the higher frequency SNPs. This result suggests that dbSNPs currently underrepresents low frequency SNPs.

The genes being resequenced by UUGC include many cell cycle and DNA repair genes. Weiss presented data on allele and haplotype frequency analyses for cyclins and cyclin-dependent kinase genes. Frequency distribution differs for each of the four cyclins g1, a2, b1 and b2. For example, low frequency SNPs are more abundant than high frequency SNPs in the cyclin g1 gene . In contrast, the frequency distribution is fairly flat or slightly U-shaped for cyclin b2. Haplotype frequency can be examined in a similar manner. For the cyclins, most haplotypes are of intermediate frequency. It is rare to observe a single dominant haplotype. Similar patterns were observed in cyclin-dependent kinase genes. SNPs were examined in the promoter of the gene encoding the chemokine 5 receptor (CCR5), which is

a coreceptor for human immunodeficiency virus (HIV). Some evidence suggests that allelic variation in this gene correlates with susceptibility and time of progression to acquired immune deficiency syndrome (AIDS) in HIV-infected individuals. Weiss analyzed 180 chromosomes from an NIH AIDS cohort and 280 chromosomes from a worldwide sample. The overall SNP frequency in the 1.1 kb promoter region was relatively high (13 total SNPs or 1/86 bp). Five SNPs had frequencies ≥ 0.25, which is an usually high number of high frequency SNPs. All common alleles were shared by human, chimp and gorilla. Allele frequency spectra showed an abundance of high frequency alleles. This may be consistent with balancing selection due to heterozygote advantage for some variants. This suggestion was confirmed by analysis of haplotype frequency. There are 2 common haplotypes and 13 total haplotypes. Statistical analysis of allele distribution in different populations provided evidence for non-neutral selection of polymorphic alleles. Weiss concluded that the CCR5 gene may be subject to directional or balancing selection.

Weiss summarized by emphasizing that SNPs should be analyzed in both coding and non-coding regions of the genome and that haplotype analysis is critical and useful, because relatively few haplotypes are often observed. Much SNP data is accumulating; however, assessing the functional significance of different SNPs, especially those in non-coding regions, remains a major challenge.

Deborah Nickerson described human SNP analysis carried out at the University of Washington Department of Genome Sciences. Nickerson pointed out that one of the motivations for SNP discovery is to provide markers for candidate gene studies. Candidate gene studies are now being applied to complex diseases, which are associated with low penetrance genetic susceptibility factors. To do a whole genome association study for disease susceptibility genes will require 30,000 to 1 million markers; SNPs can provide those markers, because there are at least 11 million SNPs in the human genome.

SNPs are the most common form of sequence variation in the human genome. They are unambiguous (diallelic) and can be determined by automated methods in large scale population studies. The method of choice for finding SNPs is direct resequencing, which must be carried out in both coding and non-coding regions of the genome. Haplotypes link SNPs in both coding and non-coding regions, but linkage disequilibrium falls off as genetic distance increases.

In functional studies on a population level, Nickerson emphasized that haplotypes should be obtained from multiple populations to maximize insight into the functional significance of polymorphism. For example, there are 80 known SNPs in the noncoding and coding regions of the angiotensin I converting enzyme (ACE) gene. Attempts have been made to explore how SNPs and haplotypes are correlated with susceptibility to hypertension, responsiveness to anti-hypertensive therapeutics and serum ACE levels. Nickerson showed that even though the search has been narrowed to a small region of the ACE gene, it is still a challenge to determine the functional consequences of sequence variation in ACE.

Many more human SNPs will be found in the near future. Efforts should focus on coding and non-coding regions and on determining haplotype. Nickerson echoed the thought expressed by Weiss in the previous talk, that functional analyses of human DNA polymorphism remains a difficult challenge.

Session II: Ethical, Legal and Social Issues

Session II Speakers:
Chair: David C. Christiani, Harvard School of Public Health
Richard Sharp, National Institute of Environmental Health Sciences, NIH
Gary Marchant, Arizona State University College of Law
Mark A. Rothstein, Institute for Bioethics

Session II Introduction and Highlights:
David Christiani introduced and chaired this session on the ethical, legal and social implications (ELSI) relating to methods and applications of toxicogenomics. The importance of ELSI resurfaced throughout the symposium, and the foresight of the symposium organizers in placing the ELSI session early in the symposium schedule was acknowledged several times. It is apparent that ELSI issues could play a major role in determining the nature of the impact of toxicogenomics on society in the next several years.

Many toxicogenomics studies focus on understanding inherited and acquired susceptibility to disease. Christiani pointed out that susceptibility is a broad and complex concept; individual susceptibility is determined by many different variables including current exposure, past exposure, health status and genetic traits. Public health policy makers have adopted two approaches to risk assessment and disease prevention for susceptible population subgroups. The high risk approach develops interventions that primarily protect the high risk portion of the population; in contrast, the general population approach develops interventions that reduce the risk of the population as a whole. It is expected that toxicogenomics will provide new and much more precise information about gene environment interactions that cause disease susceptibility. In addition, toxicogenomics will undoubtedly be used to identify individuals or population subgroups who carry specific disease susceptibility alleles. Public health policy will face the challenge of learning how to use and prevent misuse of this information.

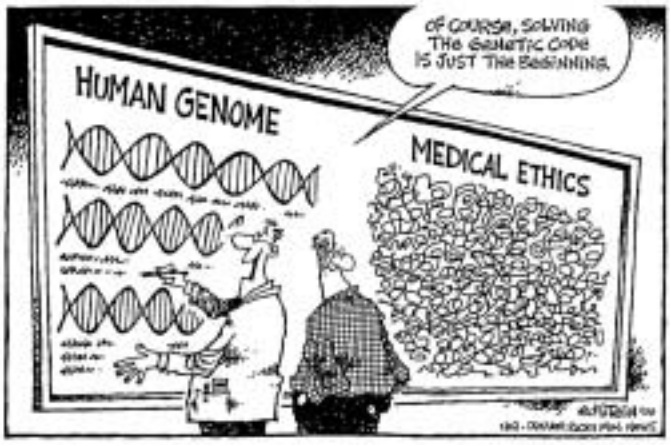

Figure 3
Ethical, Legal and Social Issues in Toxicogenomics Are As Complex or More Complex than Toxicogenomics Methods and Data

The three speakers in this session, Richard Sharp, Gary Marchant and Mark Rothstein, all emphasized that toxicogenomics has raised many ethical and legal questions, and the ELSI field is just beginning to attempt to address and resolve these questions (Figure 3). Some of the highlights of the session included the following: Sharp discussed the integrity of the relationship between research scientists and bioethicists and he challenged scientists to justify the level of "hype" given to the potential benefits of toxicogenomics research. Marchant pointed out that information on an individual's susceptibility to disease could either help or hinder the individual involved in legal disputes related to environmental exposure. Marchant also asked whether new standards are needed to estimate the significance of a change in gene expression: he raised the question, "Is a change in a microarray expression profile an 'adverse effect?'" This question was posed by many speakers in many sessions of this symposium, and it is clearly an important current concern in the field of toxicogenomics. Rothstein pointed out that many issues are unresolved concerning how genetically susceptible individuals should be treated in the workplace. For example, it is not clear whether an employer can legally deny a job to an applicant based

on his/her genetic susceptibility to disease or injury. To resolve some of these issues, Rothstein proposes that a "Bill of Rights of Examinees" should be adopted to ensure that medical testing in the workplace conforms to the expectations of an ethically appropriate doctor-patient relationship.

Session II Speaker Summaries:

Richard Sharp opened his talk by describing how the ELSI program at NIH began in 1988 under the umbrella of the Human Genome Project. Since then, the ELSI program has become an important part of the larger research program at the NHGRI, with a current annual budget of approximately $12 million. The original goals of the ELSI program were to clarify key ethical and legal issues related to the Human Genome Project, to develop policy guidelines for genomics research, to promote awareness of ELSI issues, and to support research in bioethics. In its first ten years, the priority issues were genetic discrimination by employers and insurance companies, best practices for genetic testing and informed consent, the potential psychosocial harm of genetic information or misinformation, and educational programs related to ELSI issues. The priorities of the ELSI program have shifted in the past few years towards a greater focus on human genetic variation, genetic susceptibility and/or sensitivity and preventive medicine. This shift reflects a change in the scientific focus at NHGRI. The new priorities of NHGRI and the ELSI program relate at least in part to the completion of the draft sequence of the human genome.

Sharp pointed out that the independence of the bioethicists in the ELSI program has been questioned and criticized publicly. Some critics claim that the ELSI program has served neither the interests of the general public nor those of the scholarly academic community of ethicists and philosophers. Instead, the ELSI program is viewed by these critics as a mechanism for scientists to protect themselves and to promote acceptance of new scientific technologies, primarily to benefit the scientific community itself. Sharp posed the questions: "Has the ELSI program failed to address controversial issues?; does the ELSI program simply buy 'moral credibility' for the scientific community?; and is bioethics for sale?"

Sharp answers all of these questions in the negative and he reaffirms the possibility that bioethicists can and should act as research advisors who play an integral role in development and implementation of genetic research projects involving human subjects. For example, bioethicists can clearly play a role in answering the following straight forward questions relevant to genetic research studies: what genetic research should be done on which population?; what do individuals need to know before choosing to participate as research subjects?; how should the privacy of participants be protected?; and when and in what manner should research findings be disclosed to participants? In addition, bioethicists may be able to play a significant role in answering some of the harder questions that are facing researchers in fields like toxicogenomics, as summarized below.

Researchers in toxicogenomics may want to re-examine their public image and reconsider how this field is presented in the mainstream media. Sharp posed the questions: "How should scientists present the promise and limitations of toxicogenomics research to the public? Is there too much hype [about the potential benefits of toxicogenomics]? Is the presentation [of this field to the public]sufficiently balanced?" Sharp suggested that it might be appropriate to spend an equal amount of time and effort describing both the limitations/pitfalls and the potential of this emerging technology.

Sharp also pointed out that the new techniques of toxicogenomics may for the first time begin to identify and reveal the presence of disease susceptibility genes that influence exposure outcome and health status in very large numbers of individuals in the population. Previous studies generally identified genetic traits that affect the health of small numbers of individuals. However, the consequences of carrying a disease susceptibility gene can not be predicted with high confidence, because the outcome is determined not only by a single genotype, but also by many other genetic and environmental factors. There is some danger for improper use of both toxicogenomic testing methods and the data that these methods produce (*i.e.*, to benefit employers and/or insurance companies).

As these issues are studied, bioethicists may recommend that new laws are needed to protect the rights and concerns of the individual and to ensure that individuals act in a genetically responsible manner. In the former case, employers should be required to use validated tests and to use "best practices" of informed consent, if and when they ask their employees to undergo genetic testing. On the other hand, once an individual is aware that he is at risk for a disease if exposed to a specific environmental agent, bioethicists are now asking whether that person is morally obligated to act in accordance with that knowledge. This is another area where legal and ethical issues need to be studied and clarified.

Sharp emphasized that the field of bioethics is currently in a state of transition and it is facing many complex unresolved issues. However, if bioethicists work in concert with their scientific colleagues, professionals in both of these fields and the public should reap significant benefit in the near future.

Gary Marchant discussed how toxicogenomics data on gene expression and genetic susceptibility will impact toxic tort cases and environmental regulation. In toxic tort cases, the court requires that the plaintiff provide evidence of general causation and specific causation related to his/her injury or disease. This can be a considerable burden for the plaintiff. In many cases, the evidence for general causation applies to a chemical agent or disease closely related but not identical to the agent or disease involved in the plaintiff's case (*i.e.*, cancer in a different tissue, a related but non-identical drug or chemical). An evidentiary link would be required to use this evidence, and such a link is usually not available. The availability of toxicogenomics data may make it easier for the plaintiff to win such cases, by providing additional sources of data to convince a jury of general causation. Similar arguments apply to the plaintiff's burden to prove specific causation and to provide a quantitative estimate of his/her exposure to a specific environmental agent.

As toxicogenomics research begins to identify disease susceptibility genes, the legal issues surrounding liability for exposure and its consequences will become far more complex. If an individual shows that he/she is at increased risk, this could either increase or decrease his/her chances of being compensated for damages resulting from that susceptibility. In some cases, the legal system may not support the assertion that a manufacturer must provide adequate protection to cover an individual who is more susceptible than average. However, in other cases, if it is established that genetic susceptibility lowers the baseline of exposure that causes harm, a company may be held responsible for damages suffered by the susceptible individual. The ability of consumers to mount a class action suit may also be negatively affected by the availability of genetic susceptibility data. Courts may disallow formation of a class, if it evident that the level of susceptibility, and therefore the level of risk, varies widely within members of the class. High stakes are on the table in many of these cases, and there will likely be immense pressure to use toxicigenomics-based tests before they are ready for use and/or adequately validated.

Toxicogenomics data may affect environmental regulations by reducing the level of uncertainty in risk assessment, potentially reversing the recent shift away from risk-based regulation. The new genomics-based technologies could make it easier to assess low dose effects, extrapolate risk level from animals to humans and determine mode of action of a toxicant. Personal monitoring may become feasible and allow for real-time surveillance of exposure to environmental agents. Gene expression data may also change the detection limits for "no observed effect level" or "lowest observed effect level." It will be necessary to answer the question "Is a gene expression change an adverse effect?"

Environmental regulation will also continue to struggle with balancing safety versus cost. New information on individual susceptibility to environmental pollutants also complicates precise calculation of the risk and the benefit of environmental regulation. Both toxic torts and environmental regulation will continue to evolve as toxicogenomics begins to be more widely applied in scientific, medical and legal applications.

Mark Rothstein discussed how toxicogenomics is influencing or will influence health, safety and legal issues in the workplace. Occupational issues in these areas are governed primarily by two legislative acts: the Occupational Safety and Health Act (OSHA) of 1970 and the Americans with Disabilities Act (ADA) of 1990. Rothstein outlined four main areas of concern covered by these legislations: regulation when science is uncertain, balancing the benefits and burdens of safety in the workplace; autonomy versus paternalism, and balancing the rights of employers and employees.

Many questions are emerging about how susceptible individuals should be treated in the workplace and about when genetic testing can or should be used by employers. A recent U.S. Supreme Court case affirmed the general principle that OSHA regulations apply to the average worker, and that the act is not intended to protect highly susceptible individuals to the same extent as the "normal" worker. In 1980, it was stated in writing that OSHA is not required to carry out genetic testing in a broad and general manner. Nevertheless, there are many questions about what constitutes responsible protection of the susceptible individual. Rothstein proposed that if a valid genetic test exists to establish a known genetic risk for harm in a workplace situation, then the employer should make the test available to employees without cost on a voluntary basis. In most cases, the results of the test should be available only to the employee, and the employee should then freely choose how to use knowledge of his/her genetic susceptibility. Rothstein also suggested that exceptions may exist where the risk to the employee and others is sufficiently high that the employer should also be notified of test results. It is currently unclear if employers are required to protect an individual with a greater than average susceptibility to harm in the workplace.

Questions are also developing over discrimination against individuals with genetic susceptibility to disease in the workplace. Many legal aspects of discrimination in the workplace are governed by the ADA. In this regard, Rothstein pointed out that recent rulings indicate that, within the language of the ADA, a genetically susceptible individual does not fit the definition of "disabled;" thus, it is unlikely that the genetically susceptible individual will be protected by the ADA as if he/she were disabled. There is language in ADA concerning when an employer can refuse to hire a potential employee that may be relevant to genetic susceptibility: in particular, the ADA states that an employer may refuse to hire an individual who poses a direct threat of risk to others in the workplace. At present, however, it is unclear if "others" includes the susceptible individual himself. This issue is being considered in an ongoing legal case.

Rothstein proposes that the current confusion regarding many of these issues could be resolved by enacting a "Bill of Rights of Examinees." Such a bill would ensure that medical testing in the workplace conforms to the expectations of an ethically appropriate doctor-patient relationship. For example, employees would undergo a medical test only on a voluntary basis, and would be ensured informed consent and appropriate safeguards of confidentiality if he/she submits to such a test. Recent laws were passed or are pending that address some issues surrounding genetic discrimination in the federal workplace, and 28 states have already passed laws either prohibiting genetic testing and/or genetic discrimination.

In summary, occupational health continues to evolve rapidly in response to new and anticipated genetic testing capabilities, and the impact of toxicogenomics on ethical and legal issues in occupational health is continuing to grow.

Session III: Gene Expression

Session III Speakers:
Chair: Cynthia Afshari, National Institute of Environmental Health Sciences, NIH
David Duggan, National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH
Bing Ren, University of California, San Diego
Edwin A. Clark, Millennium Pharmaceuticals, Inc.

Session III Introduction and Highlights:
Cynthia Afshari chaired and introduced the third session of the symposium on Gene Expression. Afshari pointed out that global gene expression studies could have significant impact in many fields other than toxicology. In this regard, NIEHS and the NCT want to use global transcription analyses to enhance basic and clinical research programs. The NCT is also working to build the bioinformatic structure that is needed for such research. Afshari referred to the fact that basic research scientists need to partner with statisticians and bioinformaticists to process and analyze toxicogenomics datasets. This idea was echoed during several talks in the session on Gene Expression. Highlights of this session included a study presented by David Duggan demonstrating successful use of microarray to classifying breast cancer subtypes with a high degree of accuracy; Edwin Clark used microarray data to predict the responsiveness of ovarian cancer patients to platinum-taxol therapy with an error rate of <10%; Bing Ren showed the potential of studies using promoter arrays to deduce transcriptional networks in yeast and higher eukaryotes. The importance of studying transcriptional networks to help understand pathophysiology was echoed by several speakers during the symposium.

Session III Speaker Summaries:
David Duggan described gene expression studies that define molecular signatures for different types of breast cancer. Duggan and colleagues undertook these studies because of the high public health importance of breast cancer (high incidence and second leading cause of death in American women) and because traditional histological methods are insufficient to accurately diagnose and predict breast cancer prognosis. Duggan and colleagues analyzed the expression profile of 22 tumor biopsies and 7 tumor cell lines using a cDNA microarray with 4000 known and 2500 unknown human genes. Each sample was compared to a reference sample from a non-tumorigenic immortal mammary epithelial cell line (MCF10). The data collected from these arrays was assembled into a relational database and analyzed using classical statistical analyses and five so-called "computer-intensive" algorithms.

Duggan and colleagues used the following five algorithms to analyze the tumor cell gene expression data: multidimensional scaling (MDS), hierarchical clustering dendrogram analysis, weighted gene analysis, total number of misclassifications (TNoM), and artificial neural networks (Figure 4). Duggan described the MDS and hierarchical clustering dendrogram analyses as 3- and 2-dimensional visualizations, respectively, of the set of relationships between all points (samples) in the data set. These 2 methods provide information and insight into the possible relationships within the data, but they do not evaluate its statistical significance. TNoM evaluates specific data points (genes) as discriminators between sample groups, and identifies a set of genes that are successful discriminators (minimum number of misclassifications). Weighted gene list analysis is an iterative form of MDS, in which genes are removed from the data set one at a time and the algorithm is repeated. The removal of one gene can have a small or large effect on the size of a cluster and/or its distance from other clusters. When the effects are large, the gene is likely to be a good a discriminator.

Using these methods, Duggan and colleagues identified a set of 9 genes whose gene expression profile can be used to discriminate with high accuracy between a breast cancer with a BRCA1 mutation, a BRCA2 mutation or neither (*i.e.*, sporadic breast cancer). This potentially diagnostic gene expression pattern could correctly classify 21/22 of the tumor samples used in this study. The one sample that was

misclassified was phenocopy of a BRCA1 mutant; it was genotypically wildtype but phenotypically a BRCA1 mutant, because of promoter hypermethylation and gene silencing. Thus, the gene expression profile of this tumor was correctly assessed using the set of 9 discriminator genes.

Expression Profiling Classifies BRCA1, BRCA2 and Sporadic Breast Cancer Subtypes Using Computer-Intensive Data Analyses

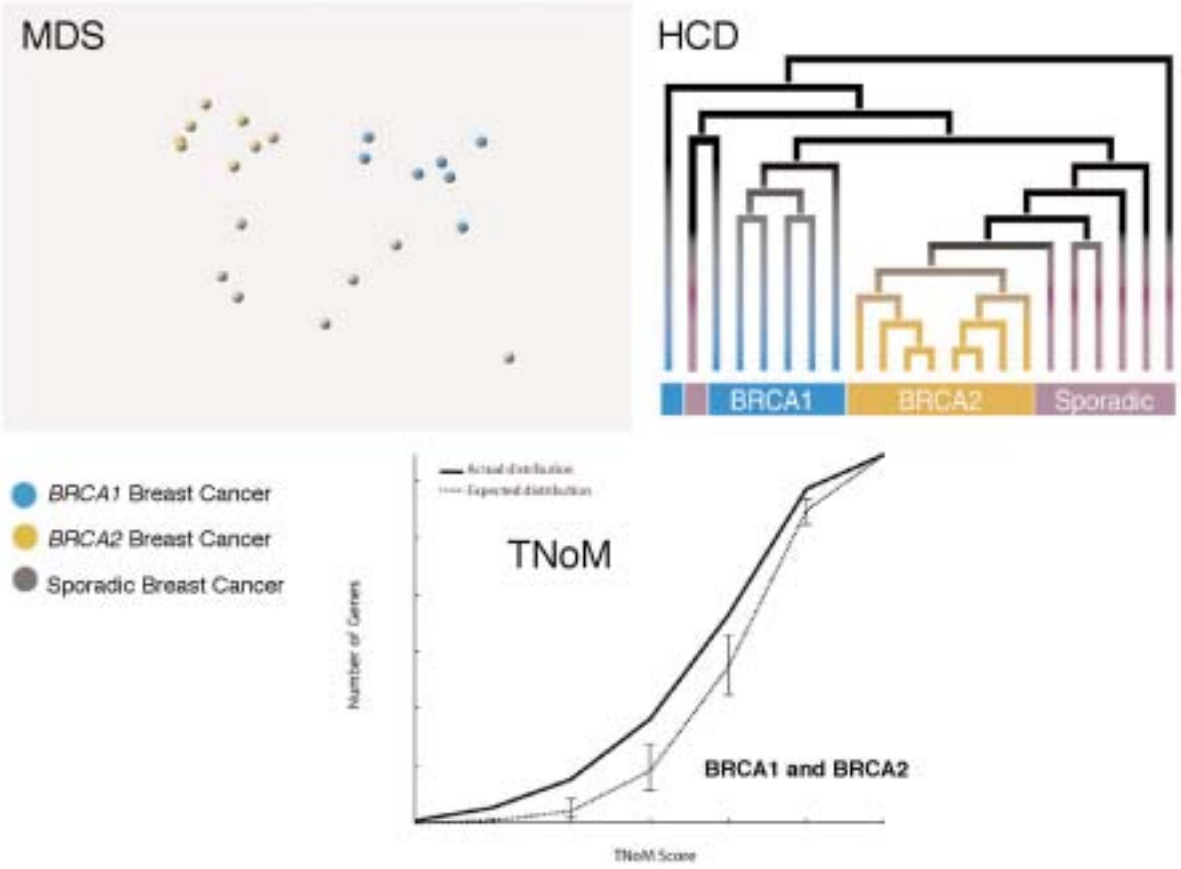

Figure 4

Expression profiling was carried out on 22 breast cancer tumor biopsies. Data was analyzed using Multidimensional Scaling (MDS), Hierarchical Clustering Dendrogram (HCD) or Total Number of Misclassifications (TNoM).

This study demonstrates that breast cancer subtypes have distinct patterns of gene expression, and that gene expression profiling is capable of classifying breast cancer subtypes with a high degree of accuracy. It is possible that this approach to studying gene expression in breast cancer tissue will improve our understanding of the molecular pathophysiology of breast cancer, and could lead to improved diagnostic and therapeutic tools.

Bing Ren developed a microarray method for identifying transcription factor targets throughout the genome of yeast or mammalian cells. Ren pointed out that transcription factors are critical players in many important cell processes. Although many eukaryotic transcription factors have been identified, it has been very difficult to identify the genes whose expression is regulated by these proteins. This is critical for our understanding of cellular signaling and the essential networks that govern complex developmental and physiological pathways.

The approach developed by Ren to study transcriptional networks is called "genome wide location analysis." A DNA sample is enriched for binding targets of the transcription factor of interest by cross-linking followed by immunoprecipitation (IP) of protein-bound DNA fragments. The IP enriched DNA sample and a control non-enriched sample are differentially labeled, mixed, and hybridized to a genomic DNA microarray that includes intergenic regions.

GWLA Identifies Gene Regulatory Networks in *Saccharomyces cerevisiae*

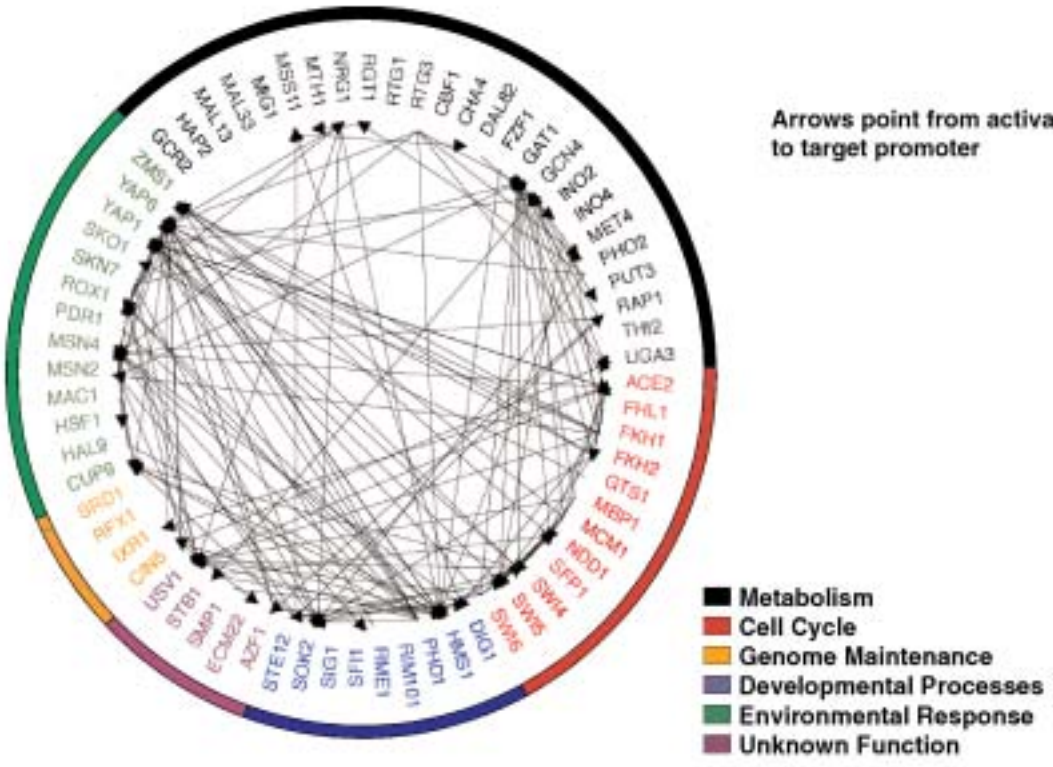

Figure 5

GWLA was initially used in yeast cells to study transcriptional targets of the well-characterized transcription factor Gal4. The results validated the method because all of the known targets of Gal4 were identified. In addition, 3 novel targets were identified that were subsequently confirmed using independent methods as functional transcriptional targets of Gal4. GWLA was then used to analyze transcriptional targets of 9 yeast cell cycle transcription factors. In an even more ambitious study, Ren and colleagues used GWLA to characterize binding targets of approximately 60 yeast transcription factors and map all their interactions (Figure 5). The data from this experiment could potentially yield information on interactions between transcriptional networks that contribute to several different metabolic pathways.

The human genome is much larger and less well annotated than the yeast genome, and human genomic DNA contains more repeated DNA sequences than yeast genomic DNA. Thus, the biggest challenge in adapting GWLA to mammalian cells in designing a microarray chip with sufficient coverage of human intergenic regions. Ren designed a microarray chip containing only promoter proximal regions (*i.e.*, 200-400 bp fragments including sequences immediately upstream and downstream of the transcription initiation site). He analyzed targets of the transcription factor E2F4 using an array with promoter regions from 1400 cell cycle regulated human genes. Amplification, labeling and hybridization conditions were optimized for specificity.

GWLA analysis showed that 10 known targets of E2F4 were enriched in the IP sample. In addition, 90 additional genes were identified as potential targets of E2F4. These genes fall into several functional clusters, including cell cycle, DNA replication, DNA repair and cell cycle checkpoints. Expression of several of the novel E2F4 target genes was also analyzed in a knockout E2F4 cell line. The results were consistent with the results of GWLA. These studies clearly demonstrate that GWLA has the potential to resolve and analyze complex transcriptional networks in lower and higher eukaryotic cells. It should have a significant impact on research on pathological mechanisms in oncogenesis and other diseases. GWLA may also yield information that leads to new diagnostic tools and therapeutic strategies to benefit human health.

Edwin Clark discussed how toxicogenomics might provide improved tools for the diagnosis and treatment of cancer. Toxicogenomics methods may prove extremely valuable in detecting early stages of disease or in predicting drug responsiveness in individual cancer patients. Especially in the case of ovarian cancer, it is thought that many deaths might be prevented if cancer could be detected earlier than is now possible. Clark explored gene profiling as a method to identify early stages of ovarian cancer.

Custom nylon microarrays were used with 12000 human genes per array. Gene expression was analyzed in approximately 90 ovarian cancer patients at all stages of disease. Profiles were also analyzed in 22 normal tissues and 50 non-ovarian cancer tissues. Clark used multiple paradigm analysis with several control/tester set combinations and two computational tools (poof and E-combo) to identify 1123 potentially important marker genes. This set of genes was eventually narrowed to 122 clusters using TP-merge and Specificity Analysis. Several interesting candidate genes including HE4 and mucin16/CA-125 were identified. Prostatin was not identified as a good diagnostic gene in this analysis. Antibodies are being made to the candidate marker proteins to clarify their usefulness as markers of ovarian cancer.

Pharmacogenomics is an analytical approach that determines whether a patient will be a responder or a non-responder when treated with a particular drug. In the case of ovarian cancer, approximately 70% of patients treated with platinum/taxane respond with complete regression, but 50% of the responders relapse within 20 months of treatment. Five year survival rate is ≈25%. Clark used a nylon array with 30000 clones to analyze cDNA expression in a discovery set (51 ovarian tumors with known drug responsiveness) and in a validation set (28 tumors with unknown drug responsiveness). Initial cluster analysis did not reveal any gene expression profiles that could distinguish responder from non-responder, and it was concluded that the clustering was driven by tumor cell type heterogeneity.

A functional genomic analysis of these data was carried out with 4 algorithms: gene cluster, linear discrimination analysis (LDA), CVSPOOF, and a supervised learning method. Two approaches were used in these analyses. The first approach was a "leave one out" cross validation approach applied to the entire discovery set of 51 samples. The resulting marker genes were tested on the 28 validation samples. This approach provided a marker set with 39% error rate in prediction of unknowns. The second approach was applied to 34 random samples from the discovery set, and tested on the remaining 17 samples in the discovery set. The algorithm was repeated 100 times, and optimized at each repetition. The markers were then tested on the validation set. This approach gave better results than the first approach, with a 21% predictive error rate. In addition, the negative predictive value for therapy had a low error rate of 10%.

In summary, Clark suggested that toxicogenomics will have impact on decision making in the clinical environment and on drug development. More rational choices can be made about when to use aggressive treatment, and therapy regimes can be individualized for a specific patient, so that new therapeutics will be used in appropriate situations. In addition, new drug targets may be identified and the drug development process may be streamlined.

Session IV: Proteomics, Biomarkers

Session IV Speakers:
Chair: James K Selkirk, National Institute of Environmental Health Sciences, NIH
Denis F. Hochstrasser, Geneva University Hospital
David R. Goodlett, Institute for Systems Biology
Emmanuel Petricoin, Food and Drug Administration

Session IV Introduction and Highlights:
James Selkirk introduced and chaired the session on Proteomics, Biomarkers. Highlights of this session included the following. Denis Hochstrasser described rapid progress in adapting proteomics to large scale applications. He introduced a new tool for protein identification and analysis, termed a molecular scanner, which employs protein electrophoresis followed by in situ trypsinization while proteins are electrophoretically transferred from a gel to a PVDF membrane. The molecular scanner has a low detection limit and high sensitivity. Emmanuel Petracoin used proteomic expression patterns to develop discriminator proteins to identify stages I-IV ovarian cancer. It is especially notable that stage I ovarian cancer was identified with 100% accuracy in a set of 503 test samples. Early detection of ovarian cancer could potentially help save many lives. Petracoin envisions a relatively rapid adoption of some proteomics tools in a clinical setting.

Session IV Speaker Summaries:
Denis Hochstrasser (Geneva University Hospital) began the session on proteomics and biomarkers with a discussion of large scale proteomics and its application in a clinical environment. In the first part of his talk, Hochstrasser described a clinical situation that demonstrates the potential impact of proteomics in a clinical situation. Hochstrasser described a 45 yr old male patient who presented at a clinic suffering from chest pain. A series of tests were performed including vital signs, electrocardiogram, chest X-ray and chest and heart exam, but these tests did not reveal obvious indications of heart disease or a high risk of coronary disease. However, a blood test indicated that the level of Troponin I (0.6 ng/mL) was approximately 2-fold higher than normal. Hochstrasser summarized studies showing that an increase in Troponin I level correlates directly with increased risk of death by heart events within 42 days. The research is sufficient to indicate that elevated Troponin level can be used as a prognostic marker for risk of death. Troponin level is also correlated with the efficacy of some therapeutic agents that prevent coronary heart obstruction. This example demonstrates the value of rapid methods that accurately detect specific proteins in the clinical environment.

Large scale proteomics are important both for rapid detection and rapid preparation of protein reagents for clinical use. GeneProt is a company based in New Jersey, with a facility in Geneva, Switzerland, that is developing large scale proteomic capacity. GeneProt has developed two platforms for large scale proteomics applications: MicroProt is designed to analyze small proteins, and Macroprot is optimized for analysis of large proteins.

In the MicroProt system, 2 HPLC systems are connected to tandem ion trap mass spectrometers (MS/MS). This arrangement avoids downtime for HPLC column maintenance (cleaning, loading, etc.) so that the analytical procedure can be operated continuously. Tandem MS/MS spectra are collected and compared with a database of theoretical spectra based on human genomic data. Data is processed simultaneously on approximately 40 units and continuously fed into computers for data storage and analysis.

The MacroProt system is based on peptide mass fingerprinting for protein identification. Samples are separated on 1 or 2-dimensional gels, proteins are extracted from the gel, digested into peptides, and then analyzed by tandem mass spectrometry. Quantitative differences between experimental and control

samples can be measured using dual fluorescent labeling technology (DIGE) or using isotope coding of different samples (*i.e.*, isotope-coded affinity tags, ICAT; see below). Robotic systems have been developed for gel processing. As for the MicroProt, samples are processed in parallel on many Macroprot systems, and data is stored and analyzed using a very large bank of computers.

GeneProt uses a laboratory information management system (LIMS) to track sample collection, protein separation and protein analysis. An automated system has also been developed for data analysis, integration and annotation.

Protein synthesis is also a major interest at GeneProt, and large capacity for rapid synthesis of useful proteins is being developed. At present, large quantities of a protein <30K can be prepared in less than one week. The goal is to have a large stock of pure proteins available, so that large scale characterization can begin as rapidly as possible once a protein of interest is identified.

Hochstrasser described a new tool for protein identification and analysis, termed a molecular scanner, developed by the Swiss Bioinformatics Institute and University of Geneva. The molecular scanner uses a protein gel in the first step of analysis. The gel is subject to shrinkage and swelling and then infused with trypsin to partially digest very large proteins. A novel hydrophilic membrane with covalently bound trypsin is then layered on the gel followed by a PVDF membrane. Proteins are electrophoretically pulsed through the trypsin saturated membrane onto the PVDF membrane, and then the PVDF membrane is placed in a mass spectrometer. Membrane-bound tryptic fragments are pushed into the MS with a laser burst and MS/MS spectra are generated. Peptide and protein identification is carried out simultaneously with protein separation. Melanie imaging software is used for visual analysis of the data and database construction proceeds automatically. The molecular scanner may prove extremely useful because it has a low detection limit and high sensitivity.

David Goodlett (The Institute for Systems Biology) described qualitative and quantitative approaches to proteomics that do not include 2D protein gels. Goodlett suggests that protein gel electrophoresis is most useful when used in a limited manner for analytical and diagnostic procedures. The non-gel based proteomic methods described by Goodlett can be used in hypothesis-driven or discovery-driven research.

Goodlett analyzed the entire proteome of the anaerobic bacterium *Haemophilus influenza* and compared the experimental results with the theoretical proteome of this organism. A four-step method was used for this study: soluble cellular protein was extracted and digested in solution; peptides were separated by capillary chromatography, peptides were fragmented and separated by tandem MS/MS. Data-dependent dynamic exclusion was not used while processing MS spectra. Instead, the MS analysis was repeated many times, each time limiting analysis to a narrow sliding window of the mass/charge spectra This approach achieves a fairly high level of proteome coverage (see below). Theoretical spectra were calculated from genomic data and compared to experimental spectra to identify expressed proteins. There are 1750 theoretical open reading frames in the *H. influenza* genome. Proteins were detected from 991 genes in bacteria grown aerobically, and from 901 genes in bacteria grown anaerobically. A total of 1255 distinct proteins were detected, which represents 74% of the *H. influenza* theoretical proteome.

Until fairly recently, most proteomic data was primarily qualitative in nature. One method for quantification of global protein expression data was recently pioneered by Rudi Aebersold of The Institute for Systems Biology. This method involves *in vitro* protein labeling using an isotope-coded cysteine affinity tag (ICAT). Tryptic peptides containing cysteine are selectively labeled, which reduces sample complexity by approximately 10-fold and achieves 90% coverage of the proteome (*i.e.*, 10% of all proteins lack cysteine). The tagged peptides are marked with biotin for affinity purification, and isotopic

label is used to differentiate experimental and control samples. The isotopically labeled peptide elutes from the mass spectrometer at a different position than the control peptide, and the relative intensity of the two peaks can be readily determined (Figure 6).

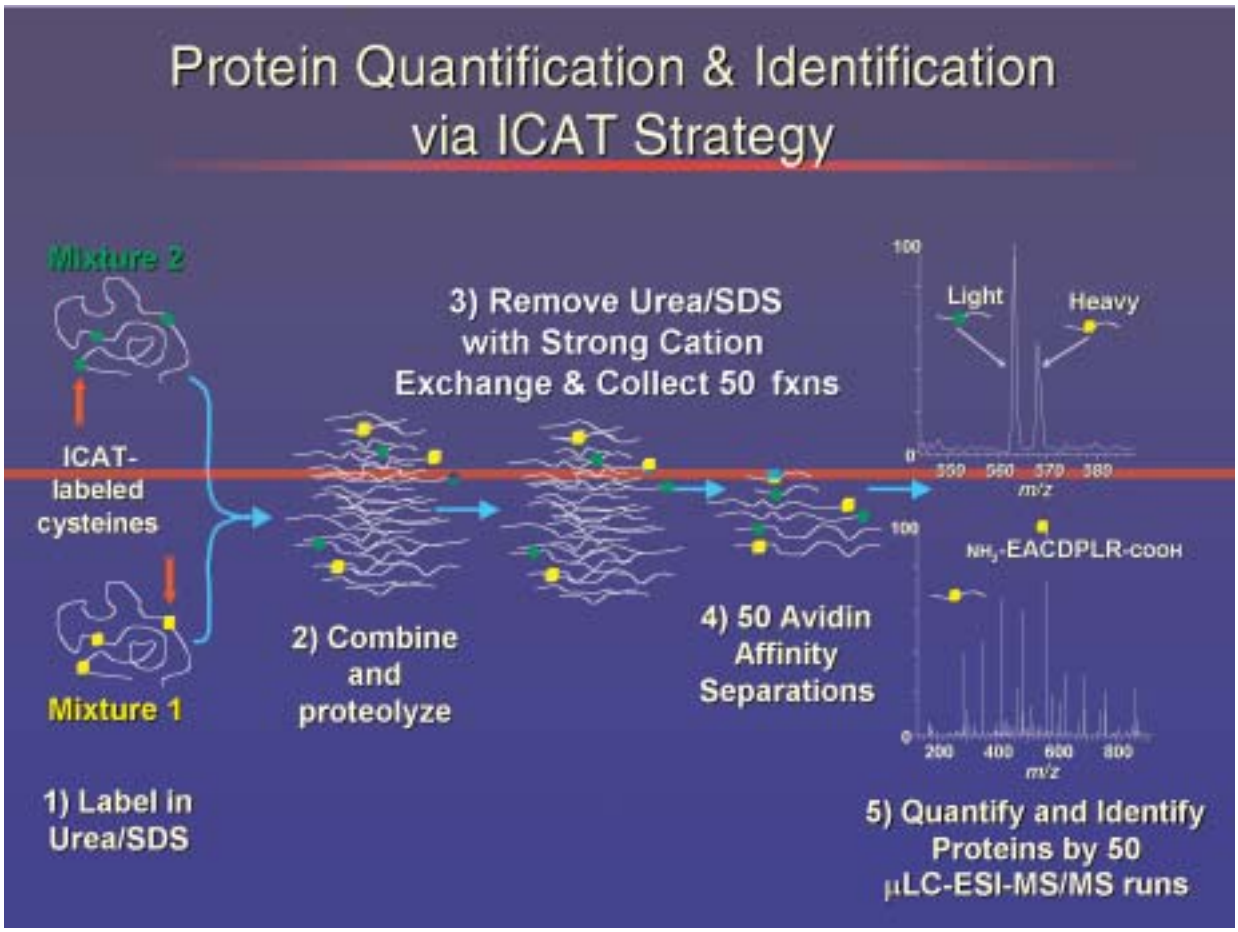

Figure 6

Goodlett used the ICAT method to study the protein composition of lipid rafts in T cells. Lipid rafts are a plasma membrane microdomain that is rich in cholesterol and glycoshingolipids and may play a role in cell signaling. Lipid rafts were isolated and analyzed in cells that were briefly stimulated with a pair of antibodies that bind to and activate the T cell receptor (TCR). The brief stimulation caused few and subtle changes in protein expression, but these changes were undetectable using protein gel electrophoresis. In contrast, change in the composition of lipid rafts was detected in TCR-stimulated cells using the ICAT method. This suggests that new proteins are recruited to the lipid rafts in response to stimulation of the TCR. The ICAT method detected some expected (*i.e.*, TCR, protein kinase C theta) and some unexpected (myosin up-regulated, vimentin down-regulated) changes in protein expression. Future goals mentioned by Goodlett include increasing throughput and achieving expression dependent analysis.

Emmanuel Petricoin (Food and Drug Administration) described a joint effort by the Food and Drug Administration and the National Cancer Institute to bring proteomic technologies to the clinic as quickly as possible. Proteomics methods are being used in ongoing clinical trials at NCI and the results are being evaluated to provide guidance in developing future proteomics applications. Petricoin described

two medium throughput and one high throughput proteomic methods that are currently being used with success.

Proteomics is important in the clinic because many therapeutic drugs exert their effects by altering the level or activity of a cellular protein target. The level or activity of the targeted protein can be measured to assess treatment efficacy. In addition, cancer is a proteomic disease, in which individual proteins or protein networks become deranged or defective. Proteomics can detect the altered proteins or pathways as a diagnostic indicator of cancer. Proteomics research is currently focusing on detecting pathogen response, profiling protein networks before and after treatment for disease, and identifying useful combination therapies.

Protein arrays are a medium throughput method that is most useful for studying a well-defined group of proteins, a network, or a specific type of protein. Sample proteins are captured on a 2-dimensional array of immobilized bait molecules such as antibodies, phage, aptamers, recombinant proteins or ligands. Sample proteins are tagged with a detection reagent that is amplified by exposure to light. Antibody are often used as bait in protein arrays. One useful approach is to array phosphospecific antibodies that query the activity of many different signaling pathways simultaneously. This method can be useful to assess efficacy of a treatment. For example, a decrease in AKT phosphorylation is indicative of a good prognosis for response to herceptin/taxol treatment in breast cancer patients.

In a reverse protein array, a sample is immobilized and queried with labeled proteins or other reagents that bind with a specific protein in the sample. Samples are titrated to ensure that the method is in the linear range of the binding curve. Reverse protein arrays are highly sensitive, linear and quantitative. Phosphoprotein antibodies are often used as probes in reverse protein arrays.

Petricoin also presented data showing that high throughput mass spectrometry has significant potential in clinical applications. In particular, Petricoin described a method for data analysis called N^{th} dimensional proteomic pattern diagnostics. This method finds buried proteomic patterns in complex MS spectra and correlates patterns with a disease state or toxicity. The algorithm finds the correct features in the complex feature set by simultaneously using supervised and unsupervised learning. N^{th} dimensional pattern diagnostics were used with an ovary training set and an ovarian cancer training set . Discriminators were developed to distinguish normal tissue, benign disorders and different stages of ovarian cancer. A test set was then examined that included 503 patients. Correct identification percentages were 94% for non-disease tissues, 97% for benign disorders, 98% for ovarian cancer stages II, III and IV, and 100% for stage I ovarian cancer. This result is important because it could potentially help avoid many cancer deaths through earlier detection of ovarian cancer.

Similar success was achieved applying N^{th} dimensional pattern diagnostics to prostate cancer tissues. Petricoin is also using this method to identify and characterize toxic responses to different types of chemicals and drugs.

In his future vision of clinical proteomics, Petricoin predicted that proteomics might be used to monitor the entire clinical process including diagnostic biopsy, choice of therapy, and post therapy assessment. Initial studies suggest that increased use of proteomics in the clinic will greatly enhance both the process and the outcome, and that the groundwork is being laid to rapidly move these methods into the clinical environment on a large scale.

Session V: Toxicogenomics

Session V Speakers:
Chair: Raymond W. Tennant, National Institute of Environmental Health Sciences, NIH
Leona D. Samson, Massachusetts Institute of Technology
Richard S. Paules, National Institute of Environmental Health Sciences, NIH
Roger R. Ulrich, Rosetta Inpharmatics, Inc.
George Orphanides, Syngenta Central Toxicology Laboratory

Session V Introduction and Highlights:
Raymond Tennant introduced and chaired the final session of the symposium on Toxicogenomics. Tennant spoke to the long term goals of the NCT, which hopes to lead progress in toxicogenomics by supporting high level research in the NIEHS intramural and extramural communities. A major part of the NCT initiative is to create, promote and maintain the Chemical Effects in Biological Systems database. A second part of the NCT is the recently established Toxicogenomics Research Consortium. Extramural participants in this Consortium are Duke University, Oregon Health Sciences University, Massachusetts Institute of Technology, Fred Hutchinson Cancer Research Center and the University of North Carolina, Chapel Hill. The NCT will promote microarray and proteomics research as well as bioinformatics and technology development to support the initiative as a whole.

Highlights of this session included the following: Richard Paules showed that microarray analyses can be used to classify toxicants, reinforcing the results of Chris Bradfield and Roger Ulrich. In addition, Paules suggested that microarray data will be useful to identify toxic endpoints such as necrosis. Ulrich demonstrated that microarray expression profiling can be used to generate mechanistic hypotheses about drug action and to evaluate the therapeutic index of a test compound. Ulrich also discussed the role of toxicogenomics in forcing a reexamination of the safety/efficacy margin for drugs. Orphanides suggested that transgenic model systems and engineered cell lines can be used to streamline the interpretation of toxicogenomics data. Orphanides also discussed the impact of toxicogenomics on risk assessment, and suggested the importance of using a comprehensive multidisciplinary toxicological approach.

Session V Speaker Summaries:
Leona Samson described research using oligo-based microarrays to analyze the cellular response to DNA alkylating agents in yeast. Exposure to DNA alkylating agents leads to altered forms of the DNA bases called adducts. Base excision repair (BER) is the primary pathway that recognizes and repairs alkylation damage in DNA. Glycosylases are important enzymes in the repair of DNA alkylation adducts, because they carry out the recognition step during adduct repair and excise the adducts from DNA. Samson's group cloned the eukaryotic 3-methyl adenine glycosylase (MAG) gene from yeast in 1980. It is a non-essential gene which is an important player in the response to reagents such as methyl methanesulfonate (MMS) and N-methyl-N'-nitro-N-nitrosoguanidine (MNNG). MAG mutants are hypersensitive to alkylating agents.

An expression profiling study was carried out in yeast to determine the global response to alkylating agents. Affymetrix chips were used that monitor expression in 6200 yeast genes. Yeast cells were treated with and without MMS, RNA isolated, cDNA prepared, differentially labeled, mixed and hybridized to arrays. A large number of inducible genes were found: 294 genes were upregulated 4- to 230-fold, and 139 genes were down-regulated significantly. In contrast, previous studies indicated that expression of a much smaller number of genes (<50) was altered in human, Chinese hamster ovary or yeast cells exposed to alkylating agents.

The MMS-inducible genes in yeast had many different cellular functions, which included the following: stress response, DNA repair, DNA replication, cell cycle, cell signaling, cell wall, cell transport, sulfur metabolism, mRNA metabolism, RNA transcription, protein secretion, cytoskeleton, chromatin structure, amino acid metabolism, protein degradation and others, including genes of unknown function. The latter category included a large proportion of the MMS-induced genes. Genes involved in amino acid metabolism and protein degradation were apparently disproportionately induced in yeast cells treated with MMS. The MMS-repressible genes also included many genes involved in protein metabolism, especially genes for ribosomal proteins.

Additional parameters were examined that influence the expression profile in these cells. A total of 26 different experimental conditions were tested, such as MMS dose, kinetics of transcriptional response, cell cycle at time of MMS treatment, and other alkylating or DNA damaging agents including MNNG, t-butyl hydrogen peroxide, γ-irradiation, 1,3 Bis[2-chloroethyl]-1-nitrosourea and 4-nitroquinoline 1-oxide. Twenty-one genes were found that respond similarly to all agents tested. These data were subject to cluster analysis and 18 "self-organized maps" or clusters were defined. One of these clusters includes yeast MAG glycosylase and 212 other genes.

The cluster containing MAG includes many genes involved in protein degradation. All of these genes carry an upstream regulatory sequence (URS2) with the sequence GGTGGCGA, which is similar to the proteosome associated control element GGTGGCAA. Furthermore, the yeast protein RPN4 binds this sequence and regulates expression from these genes. RPN4 is a protein degradation gene and a transcriptional activator. RPN4 deficient (Δrpn4) yeast are sensitive to DNA damage induced by MMS and other DNA damaging agents. In addition, expression analysis of a Δrpn4 strain indicates a very different pattern than wild type yeast in the absence and presence of MMS. Thus, these experiments describe the use of expression profiling to identify a novel and potentially important regulator of the stress response in yeast.

Samson also described a large scale discovery project to identify genes involved in the response to DNA damage in yeast. Her approach is called genomic phenotyping, and it involves screening macroarrays of yeast deletion strains for sensitivity to DNA damage. The strains are available from the Yeast Consortium and most strains can be propagated as haploid cells, but some strains must be grown as diploid. Strains are scored for growth in the presence of different doses of MMS, t-butylhydroperoxide, 4-nitroquinoline or UV. To date, 448 MMS-sensitive, 87 t-butylhydroperoxide-sensitive and 186 4-nitroquinoline-sensitive strains have been identified by screening 4800 strains. Only 4 strains were identified that are sensitive to all 4 reagents. Resistant strains were also identified. In subsequent studies, interactions between genes will be identified and the DNA repair networks will be reconstructed.

Richard Paules discussed microarray-based toxicological studies designed to assess toxic responses in rat liver. Expression profiles were analyzed in animals dosed with known toxicants to develop discriminators for different classes of chemicals. Animals were dosed with three peroxisome proliferators (clofibrate, Wyeth 14-643, and gemfibrozil), the barbiturate phenobarbital or the negative control d-mannitol. Pooled liver samples were analyzed from control animals and individual liver samples were analyzed from treated animals. All experiments were carried out in triplicate.

The expression profiles had distinct patterns for the three chemical classes, and cluster analysis identified expression patterns consistent with known mechanisms of action of the compounds. Specific gene expression patterns for animals dosed with phenobarbital were compared with information in published studies on this compound; in all cases, the observed results were consistent with published studies. However, expression profiling identified many novel genes whose expression was induced or repressed in tissues exposed to phenobarbital.

The expression pattern from animals dosed with the peroxisome proliferators or phenobarbital were analyzed extensively using several methods including a genetic algorithm/K-nearest neighbor, linear discrimination analysis, principle component analysis and analysis of variance (ANOVA). Combinations of genes were identified whose expression patterns accurately discriminated between exposure to the two classes of compounds. Gene expression profiles were also analyzed for 23 unknown samples. Twelve of the 23 unknown samples were correctly identified as similar to samples from animals exposed to peroxisome proliferators or barbiturates. In addition, 10 samples were correctly identified as not similar to peroxisome proliferator exposed samples.

Paules also used microarray expression profiling to characterize liver toxicity in rats treated with methapyriline (Figure 7). Gene expression was analyzed in animals treated with 3 doses of methapyriline for 1, 3, or 7 days. Histological analysis distinguished 4 levels of response: necrosis, bile duct

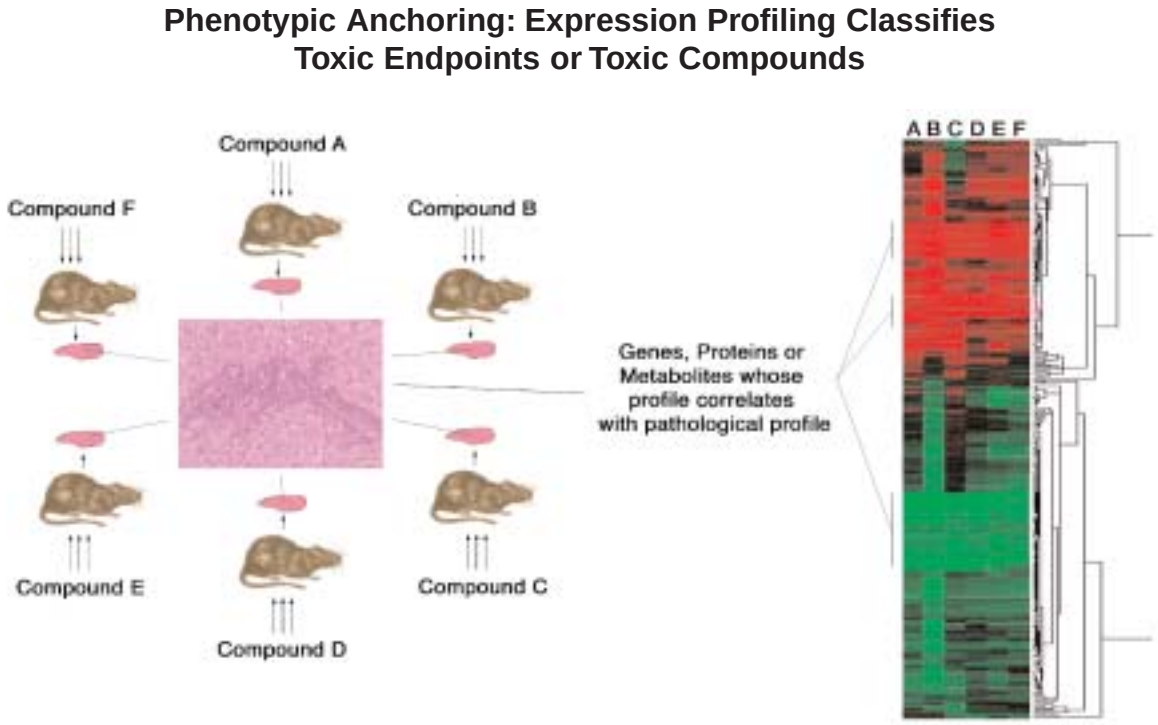

Figure 7

hyperplasia, infiltrate and microvesicular vacuolization. Expression profiling distinguished the higher doses, and the necrotic response, from the lower doses, which did not cause necrosis.
These studies show that microarray analysis can be used to classify toxicants and to identify toxic endpoints. Both types of information are extremely useful in understanding how environmental agents interact with the genome and create toxic and pathological responses in animals and humans.

Roger Ulrich described how toxicogenomics is impacting the pharmaceutical industry. There is significant pressure to produce drugs "better, faster, cheaper," and to do it with less starting material. One way to streamline drug development is to evaluate drug potential at an earlier stage in the process. A key parameter used to evaluate drugs is the therapeutic index (median toxic dose/median effective dose), which measures both beneficial and adverse effects of a compound. Ideally, toxicogenomic approaches could be used to measure therapeutic index by providing data on toxic and beneficial drug effects.

Liver toxicity is the most common type of adverse drug effect. As a conceptual test study, Ulrich used oligonucleotide-based microarrays covering 850 genes involved in toxic response pathways to study gene expression in rats dosed with 15 different hepatotoxins. The expression of a majority of the genes on the array did not change after drug treatment. However, several gene clusters were identified in animals treated with specific groups of compounds, and a small database of response profiles was generated. This approach was repeated with a larger group of 50 hepatotoxins and a 25K custom microarray with a more comprehensive set of toxic response genes in the rat. Probes for the chip were selected from a cDNA library enriched for genes that are induced or repressed in the liver of rats treated with hepatotoxins. All of these genes are ESTs and may contribute to a toxic response in rat liver, but many of them are unknown genes that completely lack annotation. The expression profiles from the 25K microarray were subject to cluster analysis and gene clusters identified according to their responsiveness to different classes of hepatotoxins. Approximately 2000/25000 genes were up- or down-regulated in response to at least one compound. Expression profiles were studied and entered into a database for evaluating additional chemicals and drugs. This database is being used for mechanistic toxicology studies. In addition, drugs are being screened with the 25K rat chip, to assess the potential of this method for assessing drug therapeutic indices.

Ulrich described two studies using these "tox chips" and liver toxicology databases. The first study explored the mechanism of the toxic response to an experimental compound that inhibits tumor necrosis factor a (TNF α) induced expression of cell adhesion molecules. TNF α upregulates transcription factor NF-κB, which in turn upregulates E-selectin, ICAM-1 and VCAM-1. The desired therapeutic effect is relatively specific inhibition of these adhesion molecules. Rats were dosed with the drug for 3 days at a high and a low dose. Clinical parameters were measured, histopathology analyzed and liver gene expression assessed using rat tox-chip microarrays. Clinical tests indicated strong hepatoxicity in animals treated at the high dose. Gene expression profiles clustered the test compound with arachlor and 3-methylcholanthrene. Arachlor acts specifically through the AhR signaling pathway which upregulates cytochrome p450 genes. Microarray expression profiles showed that the experiment compound also upregulates p450 CYP1a1 approximately 100-fold.

This information was used to generate a plausible hypothesis to explain the toxic effects of the experimental compound. NF-κB binds to AhR and sequesters it in the cytoplasm, which in effect shuts off AhR-dependent genes such as cyp1a1. Two possible modes of action were proposed for the experimental compound: 1) direct interaction with and inhibition of NF-κB; or 2) binding to AhR in a manner that inhibits formation of an AhR- NF-κB complex. These mechanisms account for down regulation of cell adhesion molecules and simultaneous upregulation of AhR-responsive genes such as cyp1a1. This testable hypothesis was generated in a relatively short time frame, and the data may be of value in the drug development process.

The second study described by Ulrich explored the efficacy and toxicity of an antisense oligonucleotide targeting protein tyrosine phosphatase 1b (PTP1b), a protein that may play a role in type 2 diabetes. PTP1b dephosphorylates and inactivates the insulin receptor (IR), down-regulating IR-dependent pathways including glucose transport, glycogen synthesis and protein synthesis. It may be possible to stimulate these IR-dependent pathways with the antisense oligonucleotide and restore insulin responsiveness in diabetic animals or patients. Expression profiling was used to measure expression of IR-responsive genes in animals treated with the oligonucleotide. Potential toxic responses, such as MAP-kinase mediated cell proliferation, were also evaluated. When obese leptin-deficient mice were dosed with the PTP1b inhibitor, insulin sensitivity was restored, genes involved in glucose metabolism were appropriately induced, and undesirable induction of genes involved in cell proliferation were not observed at low doses of the drug.

In this presentation, Ulrich demonstrated that microarray expression profiling can be used to generate mechanistic hypotheses about drug action and to evaluate the therapeutic index of a novel compound

(Figure 8). These studies are causing the pharmaceutical industry to begin to redefine toxicity indicators and re-examine the "safety margin," or the balance between safety and efficacy. It is unrealistic to expect any compound to be non-toxic at all doses under all conditions. Therefore, it is necessary to decide where to draw the line between "unsafe" and "safe." Toxicogenomics will likely have a significant impact on how this decision is made as new toxicogenomics methods and data continue to emerge. This is an important new challenge to the pharmaceutical industry and the toxicology community.

High-Throughput Global Transcript Profiling Generates Hypotheses for Low-Throughput Single Gene Studies

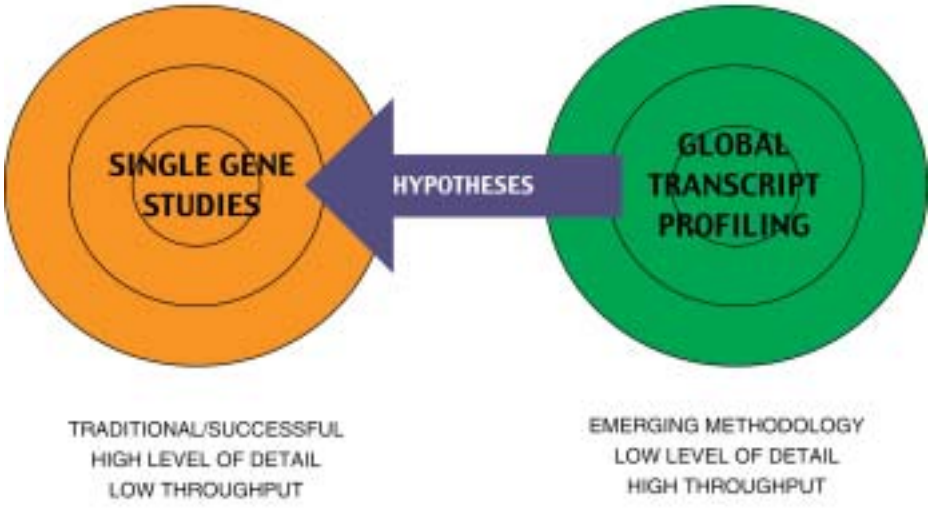

Figure 8

George Orphanides presented several approaches to streamlining the interpretation of toxicogenomics data. At present, the rate at which complex toxicogenomics data sets can be analyzed lags far behind the rate at which they can be generated. Orphanides suggested that experimental design can be used to help resolve this problem. For example, if expression profiling is carried out with a genetically altered model organism, such as a transgenic mouse or engineered cell line, the data produced will have lower complexity and higher specificity than if a wild type organism were used.

As an example of this approach, Orphanides described expression profiling in a mouse knockout strain deficient in peroxisome proliferator activated receptor α (PPAR α). PPAR α is required for cell signaling in response to many hepatocarcinogens. The activated PPAR a receptor forms a heterodimer with the retinoid X receptor and this complex transcriptionally activates many genes, some of which may mediate toxicity. Gene profiling was carried out in PPAR-proficient and PPAR–deficient animals dosed with a hepatocarcinogen, and the transcriptional effects compared. The rationale is that genes induced in cells that lack the PPAR α receptor are less likely to play a role in hepatocarcinogenesis. This approach identified a relatively small group of candidate genes, including cyp4a1, lactoferrin, ApoA-IV and others, whose role in hepatocarcinogenesis could then be characterized in detail. Orphanides currently believes that lactoferrin may play an important role in the PPAR-α dependent response to hepatocarcinogens.

Engineered cell lines can also be valuable experimental models for expression profiling. The estrogen receptor (ER) mediates cellular response to estrogenic compounds including environmental endocrine disruptors. ER α is fairly well characterized and many of its gene targets have been identified. ER β is thought to play a role in some toxic responses, but in contrast to ER α, targets of ER β are poorly characterized at present. These targets could be hard to identify in a cell that expresses both ER isoforms. Some cells express only ER α (MCF-7 breast cells) or ER β (LNCAP prostate), respectively. But these cells are not isogenic, and it is difficult to compare their expression profiles. To identify ER β - responsive genes, Orphanides obtained a cell line that does not express ER α or β, and then engineered it to produce only one ER isoform or the other. Expression profiles were compared in these

two nearly identical cell lines and the ER null cell line. Distinct genes were induced in an ER-dependent manner in these cells and target genes were identified that are induced only by ER α, only ER β, or by either receptor isoform. A similar approach is now being used to identify target genes of many nuclear orphan receptors and to identify the role played by different signaling pathways in response to hepatotoxins. In a related approach, Orphanides is also using small molecule inhibitors to selectively inactivate a specific gene or pathway of interest during expression profiling experiments. Using this approach, expression targets or signaling pathways that play a role in a specific toxic response can be selectively identified.

It is expected that toxicogenomics will have large impact on risk assessment. Because risk assessment is based on cautionary principles and safety factors that account for the large margin of error for quantitative risk of hazard, many agents with therapeutic potential are disallowed from use at their efficacious dose. Toxicogenomic data can change this by providing more mechanistic information, better extrapolation between animal and human, and identification of specific human susceptibility factors. Orphanides cautioned toxicogenomics should be used in the context of a comprehensive, multidisciplinary toxicological approach. There is danger of both misinterpreting and overinterpreting toxicogenomics data, especially while these technologies and this field are still young and developing.

This page intentionally left blank

Conference Agenda

MONDAY, DECEMBER 3

Welcome
Leona D. Samson, Massachusetts Institute of Technology

Opening Remarks
Kenneth Olden, National Institute of Environmental Health Sciences, NIH

Keynote Address
Leroy E. Hood, The Institute for Systems Biology

Special Lecture in Environmental Genomics
Michael Karin, University of California, San Diego

SESSION I: ENVIRONMENTAL GENOMICS
Chair: Daniel W. Nebert, Center for Environmental Genetics,
University of Cincinnati Medical Center

Classifying and Understanding Chemical Toxicants Using DNA Microarray Technologies
Chris A. Bradfield, McArdle Laboratory for Cancer Research

SNP Discovery in Human Cell Cycle Genes - Patterns of Diversity
Robert B. Weiss, University of Utah

SNPing in the Human Genome
Deborah A. Nickerson, University of Washington School of Medicine

Panel Discussion (Drs. Bradfield, Nebert, Nickerson, and Weiss)

SESSION II: ETHICAL, LEGAL AND SOCIAL ISSUES
Chair: David C. Christiani, Harvard School of Public Health

Ethical Issues in Genomic Research
Richard Sharp, National Institute of Environmental Health Sciences, NIH

Legal Implications of Toxicogenomics for Environmental Regulation Toxic Torts
Gary Marchant, Arizona State University College of Law

Toxicogenomics and the Workplace
Mark A. Rothstein, Institute for Bioethics

Panel Discussion (Drs. Christiani, Marchant, Denise Robinson, Rothstein and Sharp)

SESSION III: GENE EXPRESSION
Chair: Cynthia Afshari, National Institute of Environmental Health Sciences, NIH

Insights into Molecular Pathology Using cDNA Microarrays
David Duggan, National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH

Dissecting Gene Regulatory Networks – Lessons from Yeast
Bing Ren, University of California, San Diego

The Use of Genomics in Identifying, Treating, and Monitoring Cancer Patients
Edwin A. Clark, Millennium Pharmaceuticals, Inc.

Panel Discussion (Drs. Afshari, Clark, Duggan, Friend, Ren, and Cheryl Walker)
Reception, Bethesda Marriott, Pooks Hill

TUESDAY, DECEMBER 4, 2001

SESSION IV: PROTEOMICS, BIOMARKERS
Chair: James K. Selkirk, National Institute of Environmental Health Sciences, NIH

Large Scale Proteomics and the Clinic
Denis F. Hochstrasser, Geneva University Hospital

Proteomics without Polyacrylamide: Qualitative and Quantitative Mass Spectrometric Approaches to Proteomics
David R. Goodlett, Institute for Systems Biology

Clinical Proteomics: Application of New Proteomic Technology and Bioinformatic Tools to Cancer Diagnosis and Therapeutic Monitoring
Emmanuel Petricoin, Food and Drug Administration

Panel Discussion (Drs. Carol Giometti, Goodlett, Hochstrasser, and Petricoin)

SESSION V: TOXICOGENOMICS
Chair: Raymond W. Tennant, National Institute of Environmental Health Sciences, NIH

Complex Responses to Alkylating Agents
Leona D. Samson, Massachusetts Institute of Technology

Interrogation of Mechanisms Underlying Cellular Response to Environmental Stresses Using Global Gene Expression Analyses
Richard S. Paules, National Institute of Environmental Health Sciences, NIH

Poster Session in the Atrium, Natcher Center

Monitoring Changes in Gene Expression to Determine Biological Responses to Chemical Exposure
Roger R. Ulrich, Rosetta Inpharmatics, Inc.

Custom Microarray Platforms in Mechanistic Toxicity Research and Gene Promoter Characterization
George Orphanides, Syngenta Central Toxicology Laboratory

Panel Discussion (Drs. Afshari, Orphanides, Paules, Tennant, Ulrich)

Closing Remarks
Samuel H. Wilson, National Institute of Environmental Health Sciences, NIH

Program Participants

Cynthia A. Afshari, Ph.D.
Staff Scientist
Laboratory of Molecular Carcinogenesis
NIEHS Microarray Center
National Institute of Environmental Health Sciences
National Institutes of Health

Chris A. Bradfield, Ph.D.
Professor of Oncology
McArdle Laboratory for Cancer Research

David C. Christiani, M.D.
Professor
Department of Environmental Health
Harvard School of Public Health

Edwin A. Clark, Ph.D.
Senior Scientist
Millennium Pharmaceuticals, Inc.

David Duggan, Ph.D.
Associate Investigator
Microarray Unit, Genetics Section
National Institute of Arthritis Musculoskeletal Skin Diseases
National Institutes of Health

Carol S. Giometti, Ph.D.
Senior Biochemist
Biosciences Division
Argonne National Laboratory

David R. Goodlett, Ph.D.
Director, Proteomics
The Institute for Systems Biology

Denis F. Hochstrasser, M.D.
Laborataire Central de Chemie Clenique
Geneva University Hospital

Leroy E. Hood, M.D., Ph.D.
Director and President
Institute for Systems Biology

Michael Karin, Ph.D.
Professor
Department of Pharmacology
University of California, San Diego

Gary Marchant, J.D., Ph.D.
Associate Professor
Executive Director
Center for the Study of Law, Science & Technology
Arizona State University College of Law

Daniel W. Nebert, M.D.
Professor
Department of Environmental Health
Department of Pediatrics & Developmental Biology
Division of Human Genetics
University of Cincinnati Medical Center

Deborah A. Nickerson, Ph.D.
Associate Professor
Department of Molecular Biotechnology
University of Washington

Kenneth Olden, Ph.D.
Director
National Institute of Environmental Health Sciences
National Institutes of Health

George Orphanides, Ph.D.
Head of Receptor Biology and Genomics Research
Syngenta Central Toxicology Laboratory

Richard Paules, Ph.D.
Toxicogenomics Facilitator
Head, Growth Control and Cancer Group
Laboratory of Environmental Carcinogenesis and Mutagenesis
National Institute of Environmental Health Sciences
National Institutes of Health

Emanuel Petricoin, Ph.D.
Co-Director
FDA-NCI Clinical Proteomics Program
Division of Therapeutic Proteins
Center for Biologics Evaluation and Research
Food and Drug Administration

Program Participants

Bing Ren, Ph.D.
The Ludwig Institute for Cancer Research
Department of Cellular and Molecular Medicine
University of California, San Diego

Denise Robinson, Ph.D.
Executive Director
ILSI Health and Environmental Sciences Institute

Mark A. Rothstein, J.D.
Herbert F. Boehl Chair of Law and Medicine
Director
Institute for Bioethics, Health Policy and Law
University of Louisville, School of Medicine

Leona D. Samson, Ph.D.
Ellison American Cancer Society Professor
Bioengineering and Environmental Health
Massachusetts Institute of Technology

James K. Selkirk, Ph.D.
Deputy Director
National Center for Toxicogenomics
National Institute of Environmental Health Sciences
National Institutes of Health

Richard Sharp, Ph.D.
Biomedical Ethicist
National Institute of Environmental Health Sciences
National Institutes of Health

Raymond W. Tennant, Ph.D.
Director
National Center for Toxicogenomics
National Institute of Environmental Health Sciences
National Institutes of Health

Roger R. Ulrich, Ph.D., D.A.B.T.
Rosetta Inpharmatics, Inc.

Cheryl L. Walker, Ph.D.
Professor
Department of Carcinogenesis
Science Park – Research Division
University of Texas MD Anderson Cancer Center

Robert B. Weiss, Ph.D.
Associate Professor, Department of Human Genetics
Co-Director, The Utah Genome Center
University of Utah

Samuel H. Wilson, M.D.
Deputy Director
National Institute of Environmental Health Sciences
National Institutes of Health

Sponsors

Boehringer Ingelheim Pharmaceuticals, Inc.

GlaxoSmithKline

ILSI Health and Environmental Sciences Institute

For more information about the

National Center for Toxicogenomics

Please visit the NCT website at
http://www.niehs.nih.gov/nct

or send e-mail to nct@niehs.nih.gov

For more information about the

National Center for Toxicogenomics

Please visit the NCT website at
http://www.niehs.nih.gov/nct

or send e-mail to nct@niehs.nih.gov